NONPARAMETRIC KERNEL DENSITY ESTIMATION USING AUXILIARY INFORMATION FROM COMPLEX SURVEY DATA

By

SAYED A. MOSTAFA

Bachelor of Science in Statistics
Cairo University
Giza, Egypt
2010

Master of Science in Statistics
Oklahoma State University
Stillwater, OK
2014

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2016

NONPARAMETRIC KERNEL DENSITY ESTIMATION

USING AUXILIARY INFORMATION FROM COMPLEX

SURVEY DATA

Dissertation Approved:

Dr. Ibrahim Ahmad

Dissertation Adviser

Dr. Carla Goad

Dr. Ye Liang

Dr. Goutam Chakraborty

Outside Committee Member

ACKNOWLEDGMENTS

First of all, I would like to express my deepest appreciation and gratitude to Dr. Ibrahim Ahmad, my dissertation adviser, for his enthusiasm, encouragement, and guidance throughout the preparation of this dissertation.

I am grateful to my dissertation committee members; Dr. Carla Goad, Dr. Ye Liang and Dr. Goutam Chakraborty for their valuable suggestions and comments. I am also indebted to Dr. Joshua Habiger for his continuous support and the fruitful discussions we had about the work presented in this dissertation.

I would also like to acknowledge the support I received from all the statistics faculty members, graduate students and staff at OSU. Special thanks are due to Dr. Mark Payton for his continuous support and encouragement.

Many thanks to my parents, brothers and sisters for their emotional support throughout my whole life. I am grateful to Mr. Khaled Abdelaziz, my older brother, for the many favors he did for me during the time of my graduate studies.

Most of all, I wish to express my sincere thanks to my wife and my daughter *Jannah*. My wife has been extremely supportive and has made countless sacrifices to help me get to this point.

Name: SAYED A. MOSTAFA

Date of Degree: JULY, 2016

Title of Study: NONPARAMETRIC KERNEL DENSITY ESTIMATION USING AUXILIARY INFORMATION FROM COMPLEX SURVEY DATA

Major Field: STATISTICS

Abstract: This dissertation presents some new and serious attempts towards using auxiliary information effectively in kernel density estimation from complex survey data. Two approaches are proposed to develop new kernel density estimators that use both complete auxiliary information and sample information in the framework of complex surveys. Both approaches involve two steps: the first is a modeling step while the second uses the sample data and model fits from the modeling step to build efficient kernel density estimators for the density function, f, of the study variable Y. The main distinction between the two approaches is in the modeling step where in the first approach we directly model the relationship between the study variable Y and the auxiliary variable X using both parametric and nonparametric regression models while in the second approach we use nonparametric regression models to describe the relationship between a kernel-transformed study variable, say Z, and the auxiliary variable X. The first approach results in two model-assisted kernel density estimators for f. A third model-assisted kernel density estimator for f comes from the second approach. The three new estimators use the sampling weights to account for unequal probability sampling designs. The statistical properties of each of these estimators are studied under a combined design-model-based inference framework which accounts for both the underling model and the sampling scheme. The global error criterion, mean integrated squared error, is used to determine the optimal smoothing parameter for each of the three new estimators. Direct plug-in techniques are then used to obtain data-driven bandwidth estimators for these smoothing parameters. Using Monte Carlo simulation methods, we address the finite sample properties of the proposed estimators under different finite populations and sampling plans. Additionally, the performance of the new estimators relative to standard estimators that ignore the auxiliary information is assessed. On a somewhat independent track, the problem of estimating density and regression functions from samples of random sizes is considered. This problem is studied under the case of sampling with-replacement from finite populations. In this case, the effective sample size, i.e., the number of distinct sample units, is random. Based on the set of distinct sample units, kernel estimators for both density and regression functions are introduced and their statistical properties are investigated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Introduction and Literature Review

1.1. Introduction

Nonparametric density estimates, first introduced in 1951 by Fix and Hodges to free discriminant analysis from firm distributional assumptions, find wide applications in many areas of statistical analysis (Wand and Jones, 1995). These estimates provide ways of finding structure, such as skewness and multi-modality, in data sets without imposing any assumptions about the form of the models generating these data sets. Contrary to parametric methods for estimating probability density functions, which start by assuming that the unknown density function belongs to a specific parametric family such as Gaussian or gamma families and then proceed to estimate the unknown parameters using standard estimation methods such as the maximum likelihood method, the least squares method and the method of moments, nonparametric methods only put few smoothness conditions on the unknown density function. Therefore, nonparametric density estimation methods are very flexible compared to the parametric ones. We refer to Sheather (2004) for a list of fields where nonparametric density estimation techniques have been applied.

There exist several nonparametric techniques for estimating density functions. These techniques include, for example, the kernel method, the nearest neighbor method, the maximum penalized likelihood method, the orthogonal series method, wavelets and splines. Among these methods, the kernel method, first studied by Rosenblatt (1956) and Parzen (1962), is the most widely used density estimation method due to its simplicity and efficiency. The last six decades witnessed extensive work on Rosenblatt-Parzen kernel density estimators. Some of this work is covered by the excellent books of Wand and Jones (1995), Simonoff (1996), Bowman and Azzalini (1997), Pagan and Ullah (1999), Li and Racine (2007), Pons (2011) and Scott (2015), among others.

Major part of the work in density estimation has been done for data from independent and identically distributed (IID) samples. Wand and Jones (1995, Sec. 6.2.1) and Li and Racine (2007, Sec. 18.1) presented some work on kernel density estimation (KDE) under some cases where the independence assumption is tenuous. The independence assumption may fail in many situations such as time series data (cf. Ahmad (1981) and Tran (1990)). Both independence and identical distribution assumptions may not hold if the data is sampled from a finite population using a complex sampling design. For instance, clustering may create dependence between units in the same cluster and stratification may violate the identically distributed assumption (see Buskirk and Lohr (2005)). Sampling surveys represent a major source for obtaining data from finite populations. Density estimation using data from complex surveys is rare in the literature. The trails in this area are reviewed in Section 1.4.

In the literature of survey sampling, the term auxiliary information refers to any information available at the estimation stage beyond the information from the sample. If the auxiliary information is somehow correlated with the study variable, it is natural to think about utilizing this additional information to obtain improved estimates for the characteristics of the study variable. Motivated by these facts, the current dissertation focuses on kernel density estimation using auxiliary information from complex survey data. More specifically, we introduce ways to use—additional freely available—auxiliary information to develop efficient nonparametric density estimators using data from complex surveys. To best of our knowledge, no work has been done on the use of auxiliary information for nonparametric density estimation from complex survey data.

In this dissertation, complete univariate auxiliary information is used at the estimation stage in two different ways. The first approach, covered in Chapters 2 and 3, starts by modeling the relationship between the study variable Y and the auxiliary variable X using both parametric (Chapter 2) and nonparametric (Chapter 3) regression models. The model is fitted using the sample data available for both Y and X. Using the estimated model, the fitted values $\hat{Y}$ are obtained for both sampled and non-sampled Y's using the sampled X's and the X's available for the entire finite population, respectively. Both fitted values and sampled values for the study variable Y are then used to provide model-assisted kernel density estimates for the density function of Y. Under this approach, two estimators are proposed; the first uses a parametric linear regression model to obtain the fitted values $\hat{Y}$ (see Chapter 2) while the second obtains the fitted values through a nonparametric regression model (see Chapter 3). Interestingly, the mean of the first density estimator is the well-known generalized difference estimator of the finite population mean of Y while the mean of the second density estimator is the local polynomial estimator of the finite population mean of Y due to Breidt and Opsomer (2000). The statistical properties of the two density estimators are studied in Chapters 2 and 3 where asymptotic expressions are derived for the bias and the mean squared error (MSE) of each estimator under a combined design-model-based inference framework. The combined inference framework accounts for both the underlying model and the sampling scheme used which may produce unequal sampling probabilities. In these chapters, we derive the asymptotic distribution of each of the two proposed density estimators. Additionally, we use the global error criterion—mean integrated squared error ($MISE$)—to determine the optimal smoothing parameter for each estimator. In Sections 2.5 and 3.5, plug-in techniques are used to obtain data-driven bandwidth estimates.

Chapter 4 focuses on the second approach for using auxiliary information in estimating the density of the study variable Y. In this approach, nonparametric regression models are utilized to model the relationship between a kernel-transformed study variable, we call it Z, and the auxiliary variable X. Using the fitted model, we obtain the fitted values $\hat{Z}$ for all units in the finite population using the data available for X for the entire finite population. Again, we use these fitted values with the sampled values for Y to propose a third model-assisted kernel density estimator for the density of Y. Under the combined inference framework, we investigate the statistical properties of the third estimator in Sections 4.3 and 4.4. Additionally, we derive the optimal bandwidth formula and discuss data-dependent bandwidth selection techniques in Section 4.5.

Chapter 5 reports the results of a simulation study in which we investigate the performance of four kernel density estimators, namely, the standard estimator which does not use any auxiliary data and the three proposed estimators of Chapters 2–4.

Finally, on a somewhat independent track, we consider, in Chapter 6, the problem of estimating density and regression functions from samples of random size. Since samples with random sizes may arise from many situations, we confine our attention to the case of simple random sampling with replacement from finite populations. In this case, the effective sample is the set of distinct elements in the whole sample. Clearly, the size of the effective sample is random. In Section 6.2, we introduce a kernel density estimator based on the set of distinct elements and study its statistical properties. The problem of data-driven bandwidth selection is discussed in Section 6.3. The results of a limited simulation study are reported in Section 6.4. Nonparametric kernel regression based on the set of distinct elements is studied in Section 6.5.

Chapter 7 summarizes the main conclusions of the dissertation and lists some future work plans.

The rest of this Chapter is organized as follows. In Section 1.2, we give a brief introduction to nonparametric kernel density estimation from IID samples. Section 1.3 discusses three approaches commonly used for inference from survey data. The main contributions to density estimation from complex surveys are reviewed in Section 1.4. Section 1.5 reviews the use of auxiliary information in estimation from complex surveys. In Section 1.6, we introduce the asymptotic set-up adopted throughout this dissertation. A general setting and some notation are given in Section 1.7.

1.2. Kernel Density Estimation for IID Data

In this section, we briefly describe the main idea behind kernel methods for density estimation for data drawn from infinite populations by means of simple random sampling i.e, for IID data. We restrict the discussion to the univariate case.

1.2.1. Rosenblatt-Parzen Kernel Density Estimator

Suppose $Y_1, Y_2, \ldots, Y_n$ form an IID sample from an unknown distribution $F(\cdot)$ having density function $f(\cdot)$. The unknown density f is assumed to be a smooth function of y. The main goal is to use the sample data to estimate f nonparametrically. Kernel methods offer a simple way to estimating f. The *Rosenblatt-Parzen kernel density estimator* for f at any given point y is

$$\hat{f}(y;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Y_i - y}{h}\right), \tag{1.1}$$

where K is a kernel function and h is a smoothing parameter usually referred to as the bandwidth. To ensure that the estimator in (1.1) is a true density, K is usually taken to be a probability density function. Kernels that are symmetric about zero are usually used as they simplify the derivation of the properties of kernel density estimators. The standard normal density is a common choice for K, but other densities can be used as well. Asymmetric kernels can also be used in (1.1) (e.g., Chen (2000) and Abadir and Lawford (2004)). Kernels that are not densities are also sometimes used (see Wand and Jones (1995, Sec. 2.8)). The kernel function is considered as a weighing function. It uses the distance between each sample value and the estimation point y to determine weights for the sample observations to be used in constructing $\hat{f}(y;h)$. The bandwidth h determines the smoothness of the estimator. Small values of h lead to wiggly estimates having large variance while large values for h provide over-smoothed highly biased estimates. This phenomenon is known as the bias-variance trade-off in kernel density estimation.

1.2.2. Statistical Properties of The Rosenblatt-Parzen Kernel Density Estimator

The performance of kernel density estimators of the form (1.1) can be assessed using both local and global error criteria. Local error criteria measure the estimation error when estimating the density at a single point y while global ones focus on the estimation error when estimating the density over the whole real line. Taking the loss function to be the squared error loss function—the L^2-norm—the local and global error criteria are the mean squared error (MSE) and the mean integrated squared error (MISE), respectively. These criteria are defined as follows:

$$MSE\left[\hat{f}(y;h)\right] = E\left[\hat{f}(y;h) - f(y)\right]^2, \tag{1.2}$$

$$MISE\left[\hat{f}(\cdot;h)\right] = E\left[\int_{\mathbb{R}} \left\{\hat{f}(y;h) - f(y)\right\}^2 dy\right]. \tag{1.3}$$

The following conditions are standard conditions that are always made when studying the properties of kernel density estimators: (i) the density f has a continuous square integrable second derivative f'', (ii) the kernel function K is a bounded probability density function that is symmetric about zero and has finite fourth moment and (iii) the bandwidth $h \equiv h_n$ is a sequence of positive numbers which approach zero at a rate slower than n^{-1}, i.e., $h \to 0$ and $nh \to \infty$ as $n \to \infty$ (see Wand and Jones (1995, Sec. 2.5)). Under these conditions, the MSE and the MISE of the estimator in (1.1) are given by

$$MSE\left[\hat{f}(y;h)\right] = \frac{1}{nh} d_K f(y) + \frac{1}{4} h^4 c_K^2 \left\{f''(y)\right\}^2 + o\left(h^4 + \frac{1}{nh}\right), \tag{1.4}$$

$$MISE\left[\hat{f}(\cdot;h)\right] = \frac{1}{nh} d_K + \frac{1}{4} h^4 c_K^2 d_{f''} + o\left(h^4 + \frac{1}{nh}\right), \tag{1.5}$$

where $d_u = \int u^2(y)dy$ and $c_K = \int z^2 K(z)dz$. The first and second terms on the right hand side of (1.4) are the asymptotic variance and the squared asymptotic bias of $\hat{f}(y;h)$, respectively. The crucial role of h in determining the bias-variance trade-off is apparent from the MSE formula in (1.4). Using the above formulae for the MSE and the MISE, we can derive optimal local and global bandwidth values. Global optimal bandwidths, obtained by minimizing the asymptotic MISE (AMISE) of $\hat{f}(y;h)$ with respect to h, are usually preferred to local optimal bandwidths because the global ones use the same amount of smoothing for the whole range of the variable Y while local bandwidths require calculation of different h at each estimation point. For Rosenblatt-Parzen kernel density estimators, it is easy to show that the optimal global bandwidth is

$$h_{AMISE} = \left(\frac{d_K}{c_K^2 d_{f''}} \right)^{1/5} n^{-1/5}. \tag{1.6}$$

Note that the optimal bandwidth in (1.6) is not obtainable because it involves the unknown quantity $d_{f''} = \int \{f''(y)\}^2 dy$. Therefore, we need methods for selecting the appropriate amount of smoothing for kernel density estimators. The literature on kernel density estimation contains several bandwidth selection methods. Some of these methods are described in the following Section. For a recent survey of these methods, we refer to Heidenreich et al. (2013).

1.2.3. Bandwidth Selectors

Wand and Jones (1995, Ch. 3) divide the bandwidth selectors into two classes; quick and simple bandwidth selectors and hi-tech bandwidth selectors. Two popular methods in the first class are the rule-of-thumb and the over-smoothed bandwidth selection rule due to Terrell and Scott (1985) and Terrell (1990). Examples of hi-tech bandwidth selectors include plug-in and cross-validation methods. A more recent bandwidth selection method, not covered in Wand and Jones (1995), is the kernel contrast method due to Ahmad and Ran (2004). In the following two subsections, we briefly discuss some of these methods.

1.2.3.1. Rule-of-Thumb and Direct Plug-In Methods

As we mentioned earlier, the main obstacle in obtaining the optimal bandwidth in (1.6) is the unknown quantity $d_{f''}$. A simple solution is to assume that the density $f(y)$ belongs to a parametric family of distributions and then use this assumption to compute $d_{f''}$ and hence obtain h_{AMISE} using (1.6). This method for bandwidth selection is called the rule-of-thumb. Since the assumption that $f(y)$ is a normal density with variance σ^2 is usually used in the rule-of-thumb, the method is sometimes referred to as the normal scale rule. In the latter case, it is not hard to show that $d_{f''} = 3/[8\pi^{1/2}\sigma^5]$ and, thus, when K is taken to be the standard normal kernel, the optimal global bandwidth is

$$h_{NS} = \left(\frac{4}{3}\right)^{1/5} \sigma n^{-1/5}. \tag{1.7}$$

The standard deviation σ can be estimated by the sample standard deviation s, the standardized sample interquartile range or the the minimum of the two which would be a more robust measure of spread (see Wand and Jones (1995, pg. 60)).

Another intuitive bandwidth selection method is a method that replaces $d_{f''}$ in (1.6) by a kernel estimate. This method is widely known as the direct plug-in method. Kernel estimators for density functionals such as $d_{f''}$ were introduced by Hall and Marron (1987). Kernel estimators for $d_{f''}$ depends on the choice of another bandwidth, say, b. As a common practice, first use the rule-of-thumb to obtain an estimate for the pilot bandwidth b, say b_{NS}, then use that estimate to find a kernel estimate for $d_{f''}$ and finally plug that kernel estimate in (1.6) to estimate h_{AMISE}.

1.2.3.2. Cross-Validation Methods

On contrast to direct plug-in methods which, at some stage, use a subjectively chosen pilot bandwidth as shown above, cross-validation methods provide completely data-driven bandwidth selectors. The least squares cross-validation (LSCV) method was introduced by

Rudemo (1982) and Bowman (1984), independently. Stone (1984) showed that the sequence of bandwidths produced by this technique gives consistent and asymptotically optimal density estimates. The method relies on the idea of choosing h that minimizes the *MISE* of the estimator $\hat{f}(y)$ as the optimal bandwidth to be used. Expanding the squared bracket under the integral in (1.3) gives

$$MISE\left[\hat{f}(\cdot;h)\right] = E\int_{\mathbb{R}} \hat{f}^2(y;h)dy - 2E\int_{\mathbb{R}} \hat{f}(y;h)f(y)dy + \int_{\mathbb{R}} f^2(y)dy. \tag{1.8}$$

It is clear that the last term on the right hand side of (1.8) does not depend on h. Thus, minimizing the MISE of $\hat{f}(y;h)$ is equivalent to minimizing

$$MISE\left[\hat{f}(\cdot;h)\right] - \int_{\mathbb{R}} f^2(y)dy = E\left[\int_{\mathbb{R}} \hat{f}^2(y;h)dy - 2\int_{\mathbb{R}} \hat{f}(y;h)f(y)dy\right]. \tag{1.9}$$

Since the quantity on right hand side of (1.9) is unknown because it depends on f, we minimize the following unbiased estimator of this quantity:

$$LSCV(h) = \int_{\mathbb{R}} \hat{f}^2(y;h)dy - 2\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i}(Y_i;h), \tag{1.10}$$

where $\hat{f}_{-i}(y;h) = [(n-1)h]^{-1}\sum_{j\neq i}^{n} K(h^{-1}(Y_j - y))$ is the leave-one-out kernel density estimator.

An alternative for the LSCV method is the biased cross-validation (BCV) technique due to Scott and Terrell (1987). Unlike the LSCV method which works on finding h that minimizes the MISE of the estimator, the BCV technique chooses h such that the asymptotic MISE (AMISE), i.e., the sum of the first two terms in (1.5), is minimum. The main idea behind BCV is to estimate the unknown part in the AMISE formula, $d_{f''}$, by its natural estimator $d_{\hat{f}''} \equiv d_{\hat{f}^{(2)}}$ adjusted for the bias (see Scott and Terrell (1987)):

$$\hat{d}_{f''} = d_{\hat{f}''} - \left(nh^5\right)^{-1} d_{K''}, \tag{1.11}$$

where $\hat{f}^{(r)}(y) = (nh^{r+1})^{-1}\sum_{i=1}^{n} K^{(r)}(h^{-1}(Y_i - y))$. Wand and Jones (1995, pg. 80) argue that the bandwidth sequence chosen by the BCV method is considerably more stable than the one given by the LSCV method. Scott and Terrell (1987) studied the consistency and asymptotic normality of the bandwidth sequence generated by the BCV method.

1.3. Approaches for Inference from Complex Survey Data

In survey sampling literatures, there exist three different approaches to make inference about the population of interest. Here, we give a brief overview about these approaches. Consider a finite population labeled as $U = \{1, 2, \ldots, N\}$. Several survey variables, Y, W, ... etc, are associated with each unit in the finite population. Values of these variables can be observed without error only for a sample of population units. Such samples are usually drawn from the population according to some sampling design $\mathcal{P}(\cdot)$.

The most common approach used by survey statisticians to make inference about the population based on the sample data is the design-based approach. Under this approach, the finite population of interest is assumed to be fixed—does not introduce any source of randomness—and the only source of variation is the randomness induced by the sampling design. All inferences are performed with respect to the randomization distribution that results from repeated sampling (cf. Cochran (1977) and Särndal et al. (1992)). This approach ignores any parametric structure in the superpopulation that generates the finite population. Inferences made under this approach are mainly concerned with describing the current state of the finite population. Finite population quantities such as the population total $T_U = \sum_{i \in U} Y_i$, the population mean $\bar{Y}_U = N^{-1} T_U$ and the population distribution function $F_U(y) = N^{-1} \sum_{i \in U} I(Y_i \leq y)$ with $I(A)$ equals 1 if A holds and 0 otherwise, are considered parameters to be estimated from the sample. Properties of the estimators of these quantities are studied with respect to the randomization distribution. Note that under the design-based approach it is not obvious how one can perform density estimation because the finite population distribution $F_U(\cdot)$ is discrete. This point is clarified in the following section where it is shown how one could justify density estimation under the design-based approach.

The second inference approach is the traditional model-based approach which treats the finite population as a realization from an infinite hypothetical superpopulation usually termed as *model* or *distribution*. Under this approach, the sampling design is neglected, considered non-informative, by conditioning on the realized sample and all inferences about model parameters—not finite population quantities—are carried out under the assumed distribution. In other words, models are used to draw inference on superpopulations that are more general than the fixed finite population from which the sample is obtained. In this case, the kernel density estimator in (1.1) estimates the underlying density function $f(\cdot)$ of the superpopulation. The sample data $Y_1, Y_2, \ldots, Y_n$ are considered IID random variables having common density $f(\cdot)$. This is the common inference approach in statistics. As we indicated earlier, most of the literature on density estimation uses the model-based approach.

The last approach for inference from complex surveys is the combined design-model-based approach. This approach was originated by Hartley and Sielken (1975) who suggested a framework that is basically a superpopulation framework within which the design-based inference can be defined. According to Pfeffermann (1993), who formalized the combined approach, we have to account for two sources of variability under this approach; the first comes from the fact that the finite population is a realization from a superpopulation and, hence, the units of the finite population, $Y_1, Y_2, \ldots, Y_N$, are considered independent random variables having common distribution function F that has a density function f; the second source of variability comes from the sampling process that leads to the sample $Y_1, Y_2, \ldots, Y_n$ (cf. Chaudhuri and Stenger (2005) and Bleuer and Kratina (2005)). A comprehensive review of these approaches can be found in Buskirk (1999) and Buskirk and Lohr (2005). The main advantage of the combined approach is that it provides a way to make inference about model parameters using data collected from the finite population and—at the same time—it protects against any bias that could result from unequal probability sampling schemes by accounting for the variability due to the sampling design. In this dissertation, we adopt the combined design-model-based approach because of its appropriateness for our situation as we will emphasize in Section 2.3.

1.4. Nonparametric Density Estimation from Complex Survey Data

As we indicated in the previous section, the finite population distribution function $F_U(\cdot)$ is discrete and, hence, there is no density function for the finite population. Therefore, special approaches, other than the one used for density estimation from IID data, are needed. There are two approaches for density estimation from complex survey data. One approach is to consider density estimation as asymptotic descriptive inference about a limiting sequence of finite populations; that is, the finite population distribution function $F_U(\cdot)$ is assumed to converge to a differentiable distribution function $F(\cdot)$ as $N \to \infty$. This approach is used in literature to make design-based inference about $f(y) = \partial F(y)/\partial y$. Another approach is to look at density estimation from complex survey data as analytic inference about the superpopulation from which the finite population is considered to be a realization. Note that in this second approach, the goal is to estimate the hypothetical density function $f(\cdot)$ generating the values in the finite population. Either the second approach or a combination of the two approaches was the basis for the very few contributions made in the area of kernel density estimation from complex survey data.

In Section 1.2, we briefly discussed some of the literature on kernel density estimation from IID data. The IID assumptions do not generally hold for complex survey data due to possible stratification and/or clustering. Clustering is frequently used in sampling surveys to reduce the cost of data collection. This clustering generally causes data points in the same cluster to be positively correlated. Stratified sampling is known to possess higher efficiency than simple random sampling. If the finite population is stratified, it may not be reasonable to assume that all strata have the same distribution. Some work has been done in literature to adapt nonparametric density estimation to the setting of complex surveys. Below, we give a brief review of this work.

Consider a sample $s = \{y_1, y_2, \ldots, y_n\}$ drawn from a finite population having labels $U = \{1, 2, \ldots, N\}$ using some fixed-size sampling design $\mathcal{P}(\cdot)$. The sampling design $\mathcal{P}(\cdot)$ may contain stratification, clustering or a combination of stratification and clustering as in multistage sampling. Let the first and second order inclusion probabilities induced by the sampling design be defined as $\pi_i = Pr(i \in s) = \sum_{s:i\in s} \mathcal{P}(s)$ and $\pi_{ij} = Pr(i, j \in s) = \sum_{s:i,j\in s} \mathcal{P}(s)$, respectively. The basic sampling weights are $d_i = \pi_i^{-1}$ for all $i \in U$. Bellhouse and Stafford (1999) and Buskirk (1999), independently, proposed a design-weighted kernel density estimator by incorporating the basic sampling weights into the Rosenblatt-Parzen kernel density estimator given in (1.1). Their proposed estimator takes the following form:

$$\hat{f}_w(y;h) = \frac{1}{dh} \sum_{i\in s} d_i K\left(\frac{y_i - y}{h}\right), \tag{1.12}$$

where $d = \sum_{i\in s} d_i$. Under the design-based approach, both Bellhouse and Stafford (1999) and Buskirk (1999) consider the estimator in (1.12) as an estimate of the corresponding finite population smooth $f_U(y;h) = (Nh)^{-1} \sum_{i\in U} K(h^{-1}(Y_i - y))$. In addition to the weighted estimator, Bellhouse and Stafford (1999) proposed smoothed and unsmoothed binned estimators as a way to reduce the computation burden that may arise from large samples. They considered both asymptotic descriptive inference (design-based inference) and analytic inference (model-based inference). The asymptotic theory of the design-weighted kernel density estimator has been studied thoroughly by Buskirk and Lohr (2005). Under each of the three inference approaches discussed in the previous section, design-based, model-based and combined approach, they developed the consistency and asymptotic normality of the estimator in (1.12). They also discussed the bandwidth selection problem under the model-based framework.

Kernel density estimation for survey data from clustered designs was studied by Breunig (2001). Under a purely model-based approach, he studied the properties of the following estimator:

$$\hat{f}_c(y;h) = \frac{1}{nh} \sum_{c=1}^{M} \sum_{i=1}^{n_c} K\left(\frac{y_{ic} - y}{h}\right), \tag{1.13}$$

where M is the number of clusters, n_c is the sample size from the c-th cluster and $n = \sum_{c=1}^{M} n_c$. The work of Breunig (2001) accounts for the correlation structure induced by clustering. All other design features are completely ignored. Again using a pure model-based approach, Breunig (2008) studied kernel density estimators for survey data from stratified populations. He assumed that the finite population within each stratum is large enough to be well-approximated by a continuous density $f_i(\cdot)$. The main goal then is to estimate the density $f(\cdot) = \sum_{i=1}^{H}(N_i/N)f_i(\cdot)$, where H is the number of strata and N_i is the size of the i-th stratum. A sample of size n_i units is independently drawn from the i-th stratum using simple random sampling with replacement. Assuming that it is not possible to calculate a separate density estimate in each stratum, the full sample can be used to generate the following kernel density estimator for $f(y)$:

$$\hat{f}_{STR}(y;h) = \frac{1}{h\sum\sum w_{ij}} \sum_{i=1}^{H}\sum_{j=1}^{n_i} w_{ij} K\left(\frac{y_{ij}-y}{h}\right), \tag{1.14}$$

where $w_{ij} \propto \pi_{ij} = n_i/N_i$. Breunig (2008) derived the optimal bandwidth for $\hat{f}_{STR}(\cdot;h)$. He also showed that the optimal sample allocation, in the sense that it minimizes the MISE of $\hat{f}_{STR}(\cdot;h)$, is the proportional allocation, i.e., stratified sampling proportional to stratum size.

It is noteworthy that none of the work we just reviewed on kernel density estimation from complex surveys exploits any kind of auxiliary data—covariates that might be available for the sample units or for the entire finite population at the estimation stage.

1.5. The Use of Auxiliary Information in Inference from Complex Survey Data

In sample surveys, information available at the estimation stage beyond the sample information is called auxiliary information. There are two types of auxiliary information; (i) complete auxiliary information, i.e., information available for every unit in the finite population, and (ii) information available in the form of finite population totals or means. As an example on complete auxiliary information, in some countries population registers may contain information on age and taxable income for all residents. The total number of acres planted to corn reported by the government in a specific season can serve as an example on the second type of auxiliary information. For more examples on auxiliary information of both types, we refer to Breidt and Opsomer (2000) and Fuller (2009). Complete auxiliary information can be used at both design and estimation stages while auxiliary information in the form of population aggregates can be used only at the estimation stage. For instance, estimation of of labor force characteristics or household expenditure patterns might benefit from complete auxiliary data on age and taxable income (Breidt and Opsomer (2000)).

The use of auxiliary information for estimating finite population totals and means has been extensively studied. We review few examples of the work done in this area. Consider a finite population consisting of N units labeled $U = \{1, 2, \ldots, N\}$. Let X and Y be two variables measuring two characteristics of population units. Let the study variable (the variable of main interest) be Y and treat X as an auxiliary variable correlated with Y. A sample s is drawn from U using a fixed-size sampling design $\mathcal{P}(\cdot)$. Both X and Y are observed for all sampled units. Additionally, suppose X is known for all units in the population. The target is to estimate the finite population mean $\bar{Y}_U = N^{-1}\sum_{i\in U} Y_i$. A design-unbiased estimator for $\bar{Y}_U$ is the customary Horvitz-Thompson estimator

$$\hat{\bar{Y}}_{U,HT} = N^{-1}\sum_{i\in s} d_i y_i,$$

with design-variance

$$V_{\mathcal{P}}(\hat{\bar{Y}}_{U,HT}) = N^{-2}\sum\sum_{i,j\in U}\Delta_{ij}Y_iY_j, \tag{1.15}$$

where $d_i = \pi_i^{-1}$ is the sampling weight for the i-th unit in the sample, $\pi_i = Pr(i \in s)$, $\pi_{ij} = Pr(i, j \in s)$ and $\Delta_{ij} = (\pi_{ij} - \pi_i\pi_j)/\pi_i\pi_j$ (see, Särndal et al. (1992)). Note that the estimator $\hat{\bar{Y}}_{U,HT}$ does not use the auxiliary data X at all. So, a very natural question would be; can we use the auxiliary information X to improve upon the efficiency of $\hat{\bar{Y}}_{U,HT}$? The answer is yes and the main step towards such more efficient estimators is to model the relationship between X and Y using either parametric or nonparametric models. That is, assume that, in the finite population, X and Y are connected through the model

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad i \in U, \tag{1.16}$$

where both $m(\cdot)$ and $\sigma(\cdot)$ are unspecified smooth functions and the variables ε_i are independent noise with zero mean and unit variance. Different forms for $m(\cdot)$ and $\sigma(\cdot)$ give rise to many estimators of $\bar{Y}_U$ such as;

(1) the model-based prediction estimator

$$\hat{\bar{Y}}_{U,P} = N^{-1}\{\sum_{i\in s} y_i + \sum_{i\in\bar{s}} \hat{m}(X_i)\}, \tag{1.17}$$

where $\hat{m}(\cdot)$ is a parametric or nonparametric estimate of $m(\cdot)$ and $\bar{s} = U \setminus s$,

(2) the ratio estimator

$$\hat{\bar{Y}}_{U,R} = (\bar{X}_U / \hat{\bar{X}}_{U,HT})\hat{\bar{Y}}_{U,HT} \tag{1.18}$$

which is motivated by assuming that $m(X_i) = \beta X_i$ and $\sigma^2(X_i) = X_i$ for all $i \in U$,

(3) the generalized regression estimator

$$\hat{\bar{Y}}_{U,GREG} = \hat{\bar{Y}}_{U,HT} + (\bar{X}_U - \hat{\bar{X}}_{U,HT})\hat{\beta}, \tag{1.19}$$

where $\hat{\beta} = \sum_{i\in s} d_i(x_i - \bar{x})(y_i - \bar{y}) / \sum_{i\in s} d_i(x_i - \bar{x})^2$ is the design-weighted least squares estimator of β when $m(X_i) = \alpha + \beta X_i$ and $\sigma(X_i) = 1$ for all $i \in U$, and

(4) the generalized difference estimator

$$\hat{\bar{Y}}_{U,GD} = N^{-1}\{\sum_{i\in U} \hat{m}(X_i) + \sum_{i\in s} d_i y_i - \sum_{i\in s} d_i \hat{m}(x_i)\}, \tag{1.20}$$

where again $\hat{m}(\cdot)$ is a parametric (Särndal et al. (1992)) or nonparametric (Breidt and Opsomer (2000)) estimate of $m(\cdot)$.

Estimators in the form (1.17) are called model-based estimators while those in the form

(1.20) are called model-assisted estimators. Note that model-based estimators ignore the sampling weights completely and depend on the model to predict the non-sampled values of Y. Consequently, such estimators may not be asymptotically design-unbiased and/or design-consistent. Being highly dependent on the model, model-based estimators may perform badly under model misspecifications. This is in contrast to model-assisted estimators which perform very well if the model is correctly specified and remain asymptotically design-unbiased and consistent if the model is wrong. As is seen from (1.20), model-assisted estimators use model predictions $\hat{m}(x_i)$ for the population elements and then corrects the possible design-bias in that prediction using the observed differences between the predictions and the sampled elements.

As we indicated earlier in this section, both parametric and nonparametric models can be used to effectively utilize auxiliary information at the estimation stage. Nonparametric models would help to avoid the consequences of model misspecification which may cause the estimators to perform badly. This fact motivated many authors to focus on the use of nonparametric models to develop both model-based and model-assisted estimators for finite population quantities. Dorfman (1992) used nonparametric regression, to propose the following model-based estimator for the finite population total $T_U = \sum_{i \in U} Y_i$:

$$\hat{T}_{U,D} = \sum_{i \in s} y_i + \sum_{i \in \bar{s}} \hat{m}(X_i), \tag{1.21}$$

where $\hat{m}(\cdot)$ is the Nadaraya-Watson (Nadaraya (1964) and Watson (1964)) nonparametric regression estimator of $m(\cdot)$ in the model (1.16). Using design-weighted local polynomial regression estimator for $m(\cdot)$ in the model (1.16), Breidt and Opsomer (2000) proposed a model-assisted estimator for T_U:

$$\hat{T}_{U,LP} = \sum_{i \in U} \hat{m}(X_i) + \sum_{i \in s} d_i \{y_i - \hat{m}(x_i)\}. \tag{1.22}$$

Breidt et al. (2005) proposed a similar model-assisted estimator using penalized splines in place of local polynomial regression and applied their estimator to data from the Forest

Health Monitoring Survey conducted by the U.S. Forest Service. Opsomer et al. (2007) used generalized additive models to develop model-assisted estimators from the forest resources survey when multiple auxiliary variables are present. Using nonparametric regression and neural network models, Montanari and Ranalli (2005) extended the model-calibration approach of Wu and Sitter (2001) to propose nonparametric model-calibrated estimators for T_U. The model-calibrated estimators are similar in spirit to the model-assisted ones.

The finite population cumulative distribution function (CDF) $F_U(y) = N^{-1}\sum_{i\in U} I(Y_i \le y)$ is a finite population mean of the indicator variable $I(Y \le y)$. Thus, all methods discussed above can be extended to the case of estimating $F_U(y)$ to improve on the customary Horvitz-Thompson estimator $\hat{F}_{U,HT}(y) = N^{-1}\sum_{i\in s} d_i I(y_i \le y)$. Chambers and Dunstan (1986) initiated the work in this area. Using the parametric model $Y_i = \beta X_i + \sigma(x_i)\varepsilon_i$ with $\sigma(x_i) = x_i^{1/2}$ for all $i \in U$, they suggested the following model-based estimator

$$\hat{F}_{U,CD}(y) = N^{-1}\left[\sum_{i\in s} I(y_i \le y) + \sum_{j\in\bar{s},i\in s}\sum I(\hat{\varepsilon}_i \le \{(y-\hat{\beta}X_j)/\sigma(X_j)\})\right], \tag{1.23}$$

where $\hat{\varepsilon}_i = (y_i - \hat{\beta}x_i)/\sigma(x_i)$ and $\hat{\beta} = \{\sum_{i\in s} x_i y_i/\sigma^2(x_i)\}\{\sum_{i\in s} x_i^2/\sigma^2(x_i)\}^{-1}$. Rao et al. (1990) proposed both ratio and difference type estimators for $F_U(y)$:

$$\hat{F}_{U,R}(y) = N^{-1}\left[\sum_{i\in U} d_i I(\hat{\beta}X_i \le y)\right]\left[\sum_{i\in s} d_i I(\hat{\beta}x_i \le y)\right]^{-1}\left[\sum_{i\in s} d_i I(y_i \le y)\right], \tag{1.24}$$

$$\hat{F}_{U,GD}(y) = N^{-1}\left[\sum_{i\in U} I(\hat{\beta}X_i \le y) + \sum_{i\in s} d_i I(y_i \le y) - \sum_{i\in s} d_i I(\hat{\beta}x_i \le y)\right], \tag{1.25}$$

where $\hat{\beta} = \sum_{i\in s} d_i y_i / \sum_{i\in s} d_i x_i$. The performance of the model-based estimator in (1.23) relative to the model-assisted estimator in (1.25) has been investigated by Chambers et al. (1992). Dorfman and Hall (1993) used a standard nonparametric regression model, $Y_i = m(X_i) + \varepsilon_i$ with $m(\cdot)$ being sufficiently smooth and the random errors ε_i being independent with zero mean and common variance, to develop the following nonparametric versions of the estimators in (1.23) and (1.25) under the case of simple random sampling:

$$\hat{F}_{U,CDN}(y) = N^{-1}\left[\sum_{i \in s} I(y_i \leq y) + \sum_{j \in \bar{s}} \hat{G}\{t - \hat{m}(x_j)\}\right], \tag{1.26}$$

$$\hat{F}_{U,GDN}(y) = \frac{1}{n}\sum_{i \in s} I(y_i \leq y) + \frac{1}{N}\sum_{j \in \bar{s}} \hat{G}\{t - \hat{m}(x_j)\} - \left(\frac{N-n}{nN}\right)\sum_{i \in s} \hat{G}\{t - \hat{m}_i(x_j)\}, \tag{1.27}$$

where $\hat{G}(\cdot)$ is the sample empirical distribution function of the model errors, $\hat{m}(\cdot)$ is the Nadaraya-Watson kernel regression estimate of $m(\cdot)$ and $\hat{m}_i(\cdot)$ is a leave-one-out version of $\hat{m}(\cdot)$ used in the last term to simplify the mathematical derivation. Another nonparametric analog to the generalized difference estimator in (1.25) was proposed by Johnson et al. (2008):

$$\hat{F}_{U,LL}(y) = N^{-1}\left\{\sum_{i \in U} \hat{h}(X_i) + \sum_{i \in s} d_i I(y_i \leq y) - \sum_{i \in s} d_i \hat{h}(x_i)\right\}, \tag{1.28}$$

where $\hat{h}(x)$ is the local linear estimate of $h(x) = E_{\xi}[I(Y \leq y)|X = x]$. It is clear that all three model-assisted estimators in (1.25), (1.27) and (1.28) share one drawback which is the possibility of taking values outside the interval $[0, 1]$. However, since the chances of having such bad results are very limited and the effect of such values is asymptotically negligible, these estimators are very popular and practical due to their high efficiency. Silva and Skinner (1995) proposed a post-stratification estimator for $F_U(y)$ and carried out an extensive comparative study between their estimator and many of the estimators mentioned above. Chen and Wu (2002) used the parametric model-calibrated pseudo-empirical likelihood approach of Wu (1999) to estimate $F_U(y)$. Recently, calibrated estimators, in the sense of Deville and Särndal (1992), of $F_U(y)$ were proposed by Rueda et al. (2007).

It is noteworthy that all the work described above is focused on the use of auxiliary information to increase the precision of estimators of finite population quantities; means and totals. None of this work is directed to making inference about superpopulation parameters such as the superpopulation distribution function $F(\cdot)$ or the corresponding density $f(\cdot)$.

1.6. Asymptotic Set-up

The estimation of the density function $f(\cdot)$ from complex survey data can be implemented using one of three approaches; design-based approach, model-based approach and combined approach. These approaches for inference from complex survey data were discussed in Section 1.3. Associated with these approaches are three ways of handling the asymptotics:

Asymptotic set-up for the design-based approach: Assume there is a sequence of nested finite populations $\{U_\tau\}_{\tau=1}^{\infty}$ where the finite population U_τ is of size N_τ for $\tau = 1,2,3,\cdots$. The sequence $\{U_\tau\}_{\tau=1}^{\infty}$ is such that $N_\tau \to \infty$ as $\tau \to \infty$ and the corresponding sequence of finite population distribution functions satisfy $F_{U_\tau}(y) = N_\tau^{-1} \sum_{i \in U_\tau} I(Y_{i\tau} \leq y) \to F(y)$ as $\tau \to \infty$, where $F(y)$ is a smooth function having density function $f(y) = \partial F(y)/\partial y$. A sequence of samples $\{s_\tau\}_{\tau=1}^{\infty}$ with s_τ having size n_τ for $\tau = 1,2,3,\cdots$ and $n_1 < n_2 < \cdots$, is created from the sequence of populations by a sequence of sampling designs $\{\mathcal{P}_\tau\}_{\tau=1}^{\infty}$. That is, s_1 is consisted of n_1 units from U_1, s_2 contains n_2 units drawn from U_2, and so on. All limit statements are made with respect to τ, that is, as $\tau \to \infty$. For more details on this set-up for asymptotics from samples from finite populations, see Isaki and Fuller (1982). A simple example on the sequence of populations described above is given in Fuller (2009, pg. 36).

Asymptotic set-up for the model-based approach: The N finite population units, $Y_1,\ldots,Y_N$, are assumed to be a realization of independent and identically distributed random variables having common distribution $F(\cdot)$ with density function $f(\cdot)$. Considering the sampling design non-informative, the sample units $Y_1,Y_2,\ldots,Y_n$ can be seen as a direct realization from $F(\cdot)$. In this case, limiting properties of estimators are derived as $n \to \infty$.

Asymptotic set-up for the combined approach: The combined design-model-based approach for inference from complex survey data has two stages; a model stage and a design stage, as we described in Section 1.3. At the model stage, it is assumed that there exists a sequence of nested populations U_τ of size N_τ such that $N_\tau \to \infty$ as $\tau \to \infty$; $\tau = 1,2,3,\cdots$. The finite populations are generated independently from a superpopulation having distribution function $F(\cdot)$ with density $f(\cdot)$. At the design stage, a sequence of samples s_τ of size n_τ, $\tau = 1,2,3,\cdots$, with $n_1 < n_2 < \cdots$, is created from the sequence of populations by a sequence of sampling designs $\mathcal{P}_\tau$. Further, assume that the sampling rate $n_\tau/N_\tau \to \pi \in (0,1)$ as $\tau \to \infty$ with probability one (wp1) with respect to the randomization distribution. This set-up has been used by: Isaki and Fuller (1982) in the context of parametric regression estimation from complex survey data; Bellhouse and Stafford (1999) and Buskirk and Lohr (2005) in the context of kernel density estimation from complex surveys; Harms and Duchesne (2010) for nonparametric regression estimation from complex surveys.

In the current dissertation, we use the combined design-model-based approach to make inference, in the presence of auxiliary data, about the hypothetical superpopulation density function $f(\cdot)$. We slightly modify the asymptotic set-up for the combined inference approach to incorporate auxiliary information. The modified set-up is as follows: at the model stage, we assume that there exists a sequence of nested populations U_τ of size N_τ such that $N_\tau \to \infty$ as $\tau \to \infty$; $\tau = 1,2,3,\cdots$. The finite populations $\{(X_i,Y_i), i \in U_\tau\}$ are generated independently from a superpopulation having distribution $T_{XY}(x,y)$ with joint density $t_{XY}(x,y)$ and marginal densities $g_X(x)$ and $f_Y(y)$. At the design stage, a sequence of samples s_τ of size n_τ, $\tau = 1,2,3,\cdots$, with $n_1 < n_2 < \cdots$, is created from the sequence of populations by a sequence of sampling designs $\mathcal{P}_\tau$ where both X and Y are observed in each sample. The sampling rate $n_\tau/N_\tau \to \pi \in (0,1)$ as $\tau \to \infty$ wp1 with respect to the randomization distribution. All limit statements are made as $\tau \to \infty$. The index τ will be suppressed from n_τ and N_τ for ease of notation.

1.7. General Setting

In this Section, we define a general setting for the superpopulation model, the finite population and the sampling design. This setting underlies our work throughout the next four chapters. Let $U = \{1,\ldots,i,\ldots,N\}$ represent a labeling set for a finite population of N units. Associated with each unit in the finite population are real values for a study variable Y, an auxiliary variable X and possibly a design variable W; $\{(X_i, Y_i, W_i) \ : \ i \in U\}$. Values of the auxiliary variable X are known for the entire finite population. That is, $x_1, \ldots, x_N$ are known before drawing any sample. In the finite population, (X,Y) are assumed to have a joint distribution $T_{XY}(x,y)$ with joint density $t_{XY}(x,y)$ that has marginals $g_X(x)$ and $f_Y(y)$. The joint and marginal densities are real-valued functions. The parameter of interest is $f_Y(\cdot)$. A subset s of size n is selected from U according to some fixed-size sampling design $\mathcal{P}(\cdot)$. In the sample, we observe both X and Y, that is, our sample data is $\{(x_i, y_i) : i \in s\}$. The following notations will be used throughout the remaining chapters of this dissertation.

i) **Sample Inclusion Probabilities:** the first and second order inclusion probabilities from the sampling design $\mathcal{P}(\cdot)$ are $\pi_i = Pr(i \in s)$ and $\pi_{ij} = Pr(i, j \in s)$, respectively.

ii) **Sampling Weights:** the inverse of the first order inclusion probability defines the basic sampling weight $d_i = \pi_i^{-1}$ for all $i \in s$. These weights are such that $\sum_{i \in s} d_i = N$.

iii) **Model Space:** the superpopulation ξ, from which the finite population is realized, is embedded within a probability space $(\Omega, \mathscr{F}, P_\xi)$. The random variables X, Y, W are ξ-measurable where W is a design variable that determines the sampling weights. The expectation and variance operators with respect to the model are denoted by E_ξ and V_ξ.

iv) **Design Space:** the sampling design $\mathcal{P}(\cdot)$, using which the sample is drawn from the finite population, is embedded within a probability space $(S, \mathscr{S}, P_{\mathcal{P}})$. The expectation and variance operators with respect to the sampling design, respectively, are defined as $E_{\mathcal{P}}(\cdot) = E_{\mathcal{P}}(\cdot|\mathbf{X}_U, \mathbf{Y}_U)$ and $V_{\mathcal{P}}(\cdot) = V_{\mathcal{P}}(\cdot|\mathbf{X}_U, \mathbf{Y}_U)$ with $\mathbf{X_U} = (X_1, \ldots, X_N)$ and $\mathbf{Y_U}$ is defined in the same way. Alternatively, we can write $E_{\mathcal{P}}(\cdot) = E_{\mathcal{P}}(\cdot|\omega)$ with $\omega \in \Omega$.

v) **Product Space:** Assuming that, given the design variable W, the sample selection is independent of both X and Y, the product space which couples the model and the design spaces is $(\Omega \times S, \mathscr{F} \otimes \mathscr{S}, P_C)$. Combined expectation and variance operators are denoted by E_C and V_C where, for example, $E_C(\cdot) = E_\xi \left[E_{\mathcal{P}}(\cdot|\mathbf{X}_U, \mathbf{Y}_U)\right]$. For more rigorous definitions of these probability spaces, we refer to Bleuer and Kratina (2005).

CHAPTER 2

KDE Using Auxiliary Information Via Parametric Regression Models of the Study Variable

2.1. Introduction

In this chapter, we consider the use of population univariate auxiliary information to develop model-assisted kernel density estimators for the superpopulation density function of the study variable, $f_Y(\cdot)$, through modeling the relationship between the study variable Y and the auxiliary variable X using parametric regression models. In Section 2.2, we propose a model-assisted kernel density estimator for $f_Y(\cdot)$ when the relationship between X and Y can be modeled using a linear regression model. In Sections 2.3 and 2.4, the asymptotic properties of this estimator are studied carefully under the combined design-model-based inference approach described in Sections 1.3. We deal with the bandwidth selection problem for the new estimator in Section 2.5.

2.2. Model-Assisted KDE Using Linear Regression Models for Y on X

Suppose the relationship between the study variable Y and the auxiliary variable X can be described by a parametric linear regression model of the following form:

$$Y_i = \beta X_i + \sigma(X_i)\varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{2.1}$$

where β is an unknown parameter, $\sigma^2(x) = V_\xi(Y|X = x)$ is unspecified smooth function and the variables ε_i are IID with zero mean and unit variance. Note that the model in (2.1) is a linear regression model through the origin, i.e., the intercept term is assumed to be zero without loss of generality. This model includes several cases depending on the form of the variance function $\sigma(\cdot)$. One very popular case, usually used in survey sampling, is obtained when $\sigma^2(x) = x$. The importance of this special case in survey sampling comes from the fact that it yields the celebrated ratio estimator when estimating the finite population mean (see Eq. (1.18) in Section 1.5). Another case is the standard homoscedastic regression model with $\sigma^2(X_i) = \sigma^2$ for all $i \in U$. If the data for both X and Y is available for the entire finite population, the model in (2.1) can be fitted by calculating the generalized least squares estimator of the regression coefficient β which has the form

$$\beta_U = \frac{\sum_{i\in U} X_i Y_i / \sigma^2(X_i)}{\sum_{i\in U} X_i^2 / \sigma^2(X_i)}. \tag{2.2}$$

If only the sample data is available for both X and Y, the following design-weighted generalized least squares estimator of β can be used

$$\hat{\beta} = \frac{\sum_{i\in s} d_i x_i y_i / \sigma^2(x_i)}{\sum_{i\in s} d_i x_i^2 / \sigma^2(x_i)}. \tag{2.3}$$

For the first special case mentioned above where $\sigma^2(x) = x$, β_U and $\hat{\beta}$ reduce to

$$\beta_U = \frac{\sum_{i\in U} Y_i}{\sum_{i\in U} X_i} \quad \text{and} \quad \hat{\beta} = \frac{\sum_{i\in s} d_i y_i}{\sum_{i\in s} d_i x_i}, \tag{2.4}$$

while they reduce to

$$\beta_U = \frac{\sum_{i\in U} X_i Y_i}{\sum_{i\in U} X_i^2} \quad \text{and} \quad \hat{\beta} = \frac{\sum_{i\in s} d_i x_i y_i}{\sum_{i\in s} d_i x_i^2}, \tag{2.5}$$

when the variance is homoscedastic. In what follows, we will focus on the case of homoscedastic variance.

2.2.1. Proposed Estimator

If the study variable Y is known for all units in the finite population, we can write the Rosenblatt-Parzen kernel density estimator for $f_Y(y)$ as follows:

$$f_U(y;h) = \frac{1}{N}\sum_{i\in U} K_h(Y_i - y), \tag{2.6}$$

for every $y \in \mathbb{R}$, where $K_h(u) = h^{-1}K(u/h)$. Note that we can rewrite this estimator as follows:

$$f_U(y;h) = \frac{1}{N}\left[\sum_{i\in s} K_h(y_i - y) + \sum_{i\in \bar{s}} K_h(Y_i - y)\right], \tag{2.7}$$

where $\bar{s} = U \setminus s$ is the set of non-sampled units. The second term on the right hand side of (2.7) is unknown since it contains the non-sampled Y values and Y is only observed in the sample. The main idea then is to use the available auxiliary data to predict this term. This prediction can be done by using the regression model in (2.1) to obtain the fitted values $\hat{Y}$ using the known population values for X and then use these fitted values to replace the unknown Y's in the second term of (2.7). The resulting estimator for $f_Y(y)$ is

$$\hat{f}_{ml}(y;h) = \frac{1}{N}\left[\sum_{i\in s} K_h(y_i - y) + \sum_{i\in \bar{s}} K_h(\hat{Y}_i - y)\right]. \tag{2.8}$$

The estimator $\hat{f}_{ml}(\cdot)$ is a pure model-based estimator. That is, it ignores the sampling design completely. Consequently, if the model in (2.1) is misspecified, the estimator $\hat{f}_{ml}(y;h)$ may perform badly and may have undesirable properties with respect to the randomization distribution such as being design-biased and design-inconsistent when used as an estimator for the finite population quantity $f_U(y;h)$. To overcome these issues, we propose the following model-assisted kernel density estimator for the density function $f_Y(y)$:

$$\hat{f}_{dl}(y;h) = \frac{1}{N}\left[\sum_{i\in s} d_i \left\{K_h(y_i - y) - K_h(\hat{y}_i - y)\right\} + \sum_{i\in U} K_h(\hat{Y}_i - y)\right].$$

Since $\hat{Y}_i = \hat{\beta} X_i$ under the model in (2.1), our estimator can be rewritten as follows:

$$\hat{f}_{dl}(y;h) = \frac{1}{N}\left[\sum_{i \in s} d_i \left\{K_h(y_i - y) - K_h(\hat{\beta} x_i - y)\right\} + \sum_{i \in U} K_h(\hat{\beta} X_i - y)\right], \tag{2.9}$$

where $\hat{\beta}$ is as defined in (2.5). In the sequel, unless we want to emphasize the dependence of the estimator $\hat{f}_{dl}(y;h)$ on the smoothing parameter h, we will drop h and write $\hat{f}_{dl}(y)$ for compactness of notation.

2.2.2. Main Assumptions

The following set of assumptions are required for the derivation of our results.

A.1 *(The density functions)*:

(i) The density $f_Y(\cdot)$ has a bounded second derivative that is continuous and square integrable.

(ii) The density $g_X(\cdot)$ of the covariate X is continuous and bounded away from zero.

(iii) The conditional density $t_{Y|X}(\cdot|x)$ has a bounded second partial derivative and is square integrable.

A.2 *(The kernel function K)*:

(i) The function $K(\cdot)$ is a bounded density function.

(ii) $\int zK(z)dz = 0$, $\int z^2 K(z)dz = c_K < \infty$ and $\int K^2(z)dz = d_K < \infty$.

(iii) $\int z^2 \left[K'(z)\right]^2 dz = c^*_{K'} < \infty$ and $\int \left[K'(z)\right]^2 dz = d_{K'} < \infty$, where $K'(z) = dK(z)/dz$.

A.3 *(The bandwidth h)*: $h_\tau(n_\tau, N_\tau) \equiv h_\tau \equiv h$ is such that $h_\tau \to 0$ and $n_\tau h_\tau^3 \to \infty$ as $\tau \to \infty$. For simplicity, we will tend to write h instead of $h_\tau(n_\tau, N_\tau)$ in our derivations.

A.4 *(Inclusion probabilities)*: The sampling design $\mathcal{P}(\cdot)$ is independent of the covariate X and is assumed to produce inclusion probabilities that satisfy the following:

$$\min_{i \in U_\tau} \pi_i \geq \lambda > 0, \;\; \min_{i,j \in U_\tau} \pi_{ij} \geq \lambda^* > 0, \;\; \limsup_{\tau \to \infty} n_\tau \max_{i,j \in U_\tau : i \neq j} |\pi_{ij} - \pi_i \pi_j| < \infty.$$

2.3. Properties of $\hat{f}_{dl}(y)$

We start this section by showing three general properties of the proposed estimator $\hat{f}_{dl}(y)$.

Property 1. Under any sampling design, the proposed estimator, given in (2.9), is a genuine probability density function with probability approaching 1.

Proof: First, note that

$$\int_{\mathbb{R}} \hat{f}_{dl}(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i \left\{\int_{\mathbb{R}} K_h(y_i - y)dy - \int_{\mathbb{R}} K_h(\hat{\beta}x_i - y)dy\right\} + \sum_{i\in U}\int_{\mathbb{R}} K_h(\hat{\beta}x_i - y)dy\right].$$

Using the symmetry of the kernel $K(\cdot)$ and the change of variables $z = (y-u)/h$, we have

$$\int_{\mathbb{R}} \hat{f}_{dl}(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i \left\{\int_{\mathbb{R}} K(z)dz - \int_{\mathbb{R}} K(z)dz\right\} + \sum_{i\in U}\int_{\mathbb{R}} K(z)dz\right] = 1.$$

Second, $\hat{f}_{dl}(y) \geq 0$ for all $y \in \mathbb{R}$ with probability approaching 1 since the estimator $\hat{f}_{dl}(y)$ is a consistent estimator, in the MSE sense, for $f(y)$ (see the MISE expression in Theorem 2.2). □

It is noteworthy that, like many other density estimators in literature, the estimator $\hat{f}_{dl}(y)$ can take negative values specially if the assumed working model, the linear model in (2.1) in the present case, is misspecified. A simple remedy is to define the following L^2-projection of $\hat{f}_{dl}(y)$ onto a class of non-negative densities:

$$\check{f}_{dl}(y) = max\{0, \hat{f}_{dl}(y) - c\}, \tag{2.10}$$

where c is a normalizing constant to make the estimate $\check{f}_{dl}(y)$ integrate out to 1 (see Glad et al. (2003)). This simple adjustment is a common practice in the literature of nonparametric density estimation to handle possible negativity that may occur in many nonparametric density estimators such as higher order kernel density estimators, orthogonal series density estimators, splines and wavelets (e.g., Terrell and Scott (1980, pg. 1160), Glad et al. (2003, pg. 415), Scott (2004, pg. 8) and Dassanayake et al. (2015, pg. 4)). Glad et al. (2003) have shown that the constant c in (2.10) always exists and is unique and, thus, the estimator $\check{f}(\cdot)$ is well-defined for any density estimator $\hat{f}(\cdot)$ that is bounded, square integrable and satisfies

$$\int max\{0, \hat{f}(y)\}dy \geq 1. \tag{2.11}$$

Moreover, they prove that under these conditions, the estimator $\check{f}(\cdot)$ is at least as efficient, in the sense of mean integrated squared error, as the original estimator $\hat{f}(\cdot)$. The following lemma shows that the proposed estimator $\hat{f}_{dl}(\cdot)$ satisfies the above conditions and, hence, the adjustment in (2.10) does not result in any loss in efficiency.

Lemma 2.1. *The estimator $\hat{f}_{dl}(y)$, defined in* (2.9)*, is bounded, square integrable and satisfies the inequality in* (2.11).

Proof: First, note that the estimator $\hat{f}_{dl}(y)$ can be written as follows:

$$\hat{f}_{dl}(y;h) = \frac{1}{Nh}\sum_{i\in s} d_i K\left(\frac{y-y_i}{h}\right) - \frac{1}{Nh}\sum_{i\in s} d_i K\left(\frac{y-\hat{y}_i}{h}\right) + \frac{1}{Nh}\sum_{i\in U} K\left(\frac{y-\hat{Y}_i}{h}\right). \tag{2.12}$$

Clearly, the kernel function K is square integrable by assumption A.2(ii). Consider the first term on the right hand side of (2.12) and notice that it is a linear combination of the same kernel function evaluated at different points. This term is square integrable since linear combinations of square integrable functions are also square integrable (see Howell (2001, pg. 400)). Similarly, each of the other two terms on the right hand side of (2.12) is square integrable. Therefore, the estimator $\hat{f}_{dl}(y)$ is square integrable as a function of y.

Second, to check the condition in (2.11), let $A = \{y : \hat{f}_{dl}(y;h) \geq 0\}$ and observe that from Property 1 above we have

$$\begin{aligned}
1 &= \int_{\mathbb{R}} \hat{f}_{dl}(y;h)dy \\
&= \int_A \hat{f}_{dl}(y;h)dy + \int_{A^c} \hat{f}_{dl}(y;h)dy \\
&= \int_{\mathbb{R}} max\{0, \hat{f}_{dl}(y;h)\}dy + \int_{A^c} \hat{f}_{dl}(y;h)dy \\
&\leq \int_{\mathbb{R}} max\{0, \hat{f}_{dl}(y;h)\}dy,
\end{aligned}$$

which completes the proof of the lemma. $\square$

Terrell and Scott (1980) commented on the issue of possible negativity of density estimators saying "*generally these estimates are negative only in the tails and, thus, for many applications the lack of non-negativity is unimportant*". However, the adjustment in (2.10) would be important in some applications such as using the density estimate for random number generation. For the proposed estimator, if the working model is correctly specified, such adjustment may not be needed at all because the possibility of negative values becomes almost zero in this case due to the balance that occurs between the last two terms of (2.9). Theorem 2.2 below states that the proposed estimator $\hat{f}_{dl}(y)$ is a consistent estimator for $f(y)$ and hence the possible negativity does not cause any serious problem and is asymptotically negligible. This fact plus the high efficiency of model-assisted estimators made the many non-bona fide model-assisted estimators of the finite population distribution function described in Section 1.5 maintain their practicability. In our empirical study reported in Chapter 5, we found the probability of negativity of any of the three model-assisted density estimators of Chapters 2-4 is very minor and does not exceed 0.5% when the working model is correctly specified and 1% for misspecified working models. Since, our proposed estimators cover several modeling situations, model misspecification should not be problematic. For instance, if we are not sure about the form of the model, we can resort to the estimators in Sections 3.2.1 and 4.2.1 which do not depend on any parametric model.

Property 2. Under any sampling design, the mean of the proposed estimator is the well-known generalized regression estimator of the finite population mean $\bar{Y}_U$ (see Section 1.5):

$$\hat{\bar{Y}}_{U,GREG} = \int_{\mathbb{R}} y\hat{f}_{dl}(y)dy = \hat{\bar{Y}}_{U,HT} + \hat{\beta}(\bar{X}_U - \hat{\bar{X}}_{U,HT}),$$

where $\hat{\bar{Z}}_{U,HT} = N^{-1}\sum_{i\in s} d_i Z_i$.

Proof:

$$\int_{\mathbb{R}} y\hat{f}_{dl}(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i \int_{\mathbb{R}} yK_h(y_i - y)dy - \sum_{i\in s} d_i \int_{\mathbb{R}} yK_h(\hat{\beta}x_i - y)dy\right.$$

$$+\sum_{i\in U}\int_{\mathbb{R}} yK_h(\hat{\beta}x_i-y)dy\Bigg].$$

Again, using the symmetry of $K(\cdot)$ and the change of variables $z=(y-u)/h$, we have

$$\begin{aligned}\int_{-\infty}^{\infty}\hat{f}_{dl}(y)dy &= \frac{1}{N}\Bigg[\sum_{i\in s}d_i\int_{-\infty}^{\infty}(y_i+hz)K(z)dz-\sum_{i\in s}d_i\int_{-\infty}^{\infty}(\hat{\beta}x_i+hz)K(z)dz \\ &\qquad +\sum_{i\in U}\int_{-\infty}^{\infty}(\hat{\beta}x_i+hz)K(z)dz\Bigg] \\ &= \frac{1}{N}\left[\sum_{i\in s}d_iy_i-\hat{\beta}\sum_{i\in s}d_ix_i+\hat{\beta}\sum_{i\in U}x_i\right] \\ &= \frac{1}{N}\sum_{i\in s}d_iy_i+\hat{\beta}\left[\frac{1}{N}\sum_{i\in U}x_i-\frac{1}{N}\sum_{i\in s}d_ix_i\right],\end{aligned}$$

where the second equality follows from Assumptions A.2(i–ii). □

Property 3. Integrating the estimator $\hat{f}_{dl}(y)$ over the interval $(-\infty,t]$ gives a smooth version of Rao et al. (1990) model-assisted estimator of the finite population CDF, $F_U(t)$ (see Eq. (1.25) in Section 1.5):

$$\begin{aligned}\hat{F}_{GDS}(t) &= \int_{-\infty}^{t}\hat{f}_{dl}(y)dy \\ &= \frac{1}{N}\left[\sum_{i\in U}\mathcal{K}\left(\frac{t-\hat{\beta}X_i}{h}\right)+\sum_{i\in s}d_i\mathcal{K}\left(\frac{t-y_i}{h}\right)-\sum_{i\in s}d_i\mathcal{K}\left(\frac{t-\hat{\beta}x_i}{h}\right)\right],\end{aligned}$$

where $\mathcal{K}(y)=\int_{-\infty}^{y}K(u)du$.

Proof:

$$\begin{aligned}\int_{-\infty}^{t}\hat{f}_{dl}(y)dy &= \frac{1}{N}\Bigg[\sum_{i\in s}d_i\int_{-\infty}^{t}K_h(y_i-y)dy-\sum_{i\in s}d_i\int_{-\infty}^{t}K_h(\hat{\beta}x_i-y)dy \\ &\qquad +\sum_{i\in U}\int_{-\infty}^{t}K_h(\hat{\beta}x_i-y)dy\Bigg] \\ &= \frac{1}{N}\left[\sum_{i\in s}d_i\int_{-\infty}^{\frac{t-y_i}{h}}K(u)du-\sum_{i\in s}d_i\int_{-\infty}^{\frac{t-\hat{\beta}x_i}{h}}K(v)dv+\sum_{i\in U}\int_{-\infty}^{\frac{t-\hat{\beta}X_i}{h}}K(w)dw\right] \\ &= \frac{1}{N}\left[\sum_{i\in s}d_i\mathcal{K}\left(\frac{t-y_i}{h}\right)-\sum_{i\in s}d_i\mathcal{K}\left(\frac{t-\hat{\beta}x_i}{h}\right)+\sum_{i\in U}\mathcal{K}\left(\frac{t-\hat{\beta}X_i}{h}\right)\right],\end{aligned}$$

where the last equality follows from the fact that K is a density function. $\square$

Now we study the statistical properties of the estimator in (2.9) under the combined design-model-based approach. Recall from Section 1.6 that under the combined approach, to calculate, for example, the expectation of the estimator $\hat{f}_{dl}(y)$, we first calculate the design-expectation by conditioning on the realized finite population. That is, we calculate $E_{\mathcal{P}}\left[\hat{f}_{dl}(y)|\mathbf{X}_U,\mathbf{Y}_U\right] \equiv E_{\mathcal{P}}\left[\hat{f}_{dl}(y)|\omega\right]$ with $\omega \in \Omega$. For compactness of notation, the conditioning argument will be suppressed when no confusion is expected. Then we work out the model-expectation of the resulting design-expectation.

The density estimator $\hat{f}_{dl}(\cdot)$ in (2.9) involves the estimate $\hat{\beta}$, defined in (2.3), which complicates the derivation of the asymptotic properties of the density estimator itself. We pass this issue by replacing the sample-based estimate $\hat{\beta}$ in (2.9) with the finite population estimate β_U, which is defined in (2.2), to get the following pseudo density estimator:

$$\tilde{f}_{dl}(y;h) = \frac{1}{N}\left[\sum_{i\in s} d_i\left\{K_h(y_i - y) - K_h(\beta_U x_i - y)\right\} + \sum_{i\in U} K_h(\beta_U X_i - y)\right]. \tag{2.13}$$

The estimator $\hat{f}_{dl}(\cdot)$ is a pseudo estimator because it is not obtainable as it depends on the unknown quantity β_U. Appealing to the design-consistency of Horvitz-Thompson estimators shows that $\hat{\beta}$—being a ratio of two Horvitz-Thompson estimators, see Eq. (2.5)—is design-consistent for β_U. Given this consistency, we use results from Randles (1982) to show that the two estimators $\hat{f}_{dl}(\cdot)$ and $\tilde{f}_{dl}(\cdot)$ have the same limiting design-based distribution. This result is formulated in the following lemma.

Lemma 2.2. *Under the smoothness assumptions of Section 2.2.2 on the kernel $K(\cdot)$, the estimator $\hat{f}_{dl}(y;h)$ has the same limiting design-based distribution as $\tilde{f}_{dl}(y;h)$.*

Proof: We use similar notation to that in Randles (1982). Let γ be a mathematical variable and denote $\hat{f}_{dl}(y)$ as $T_n(\hat{\beta})$ and $\tilde{f}_{dl}(y)$ as $T_n(\beta_U)$. Define the sample indicators as: $I_i = 1$ if $i \in s$ and $I_i = 0$ otherwise, for all $i \in U$. Therefore, I_i is a Bernoulli random variable with mean $E_{\mathcal{P}}(I_i) = \pi_i$, the sample inclusion probability of the i-th unit. Note that

$$\begin{aligned}
E_{\mathcal{P}}\left[T_n(\gamma)\right] &= \frac{1}{N}E_{\mathcal{P}}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i)-K_h(y-\gamma x_i)\right\}+\sum_{i\in U}K_h(y-\gamma X_i)\right] \\
&= \frac{1}{N}E_{\mathcal{P}}\left[\sum_{i\in U} I_i d_i\left\{K_h(y-Y_i)-K_h(y-\gamma X_i)\right\}+\sum_{i\in U}K_h(y-\gamma X_i)\right] \\
&= \frac{1}{N}\left[\sum_{i\in U} E_{\mathcal{P}}(I_i) d_i\left\{K_h(y-Y_i)-K_h(y-\gamma X_i)\right\}+\sum_{i\in U}K_h(y-\gamma X_i)\right] \\
&= \frac{1}{N}\left[\sum_{i\in U} \left\{K_h(y-Y_i)-K_h(y-\gamma X_i)\right\}+\sum_{i\in U}K_h(y-\gamma X_i)\right] \\
&= \frac{1}{N}\sum_{i\in U} K_h(y-Y_i) = f_U(y;h).
\end{aligned}$$

Therefore, the limiting mean function is

$$\mu(\gamma) = \lim_{n\to\infty} E_{\mathcal{P}}\left[T_n(\gamma)\right] = f(y), \tag{2.14}$$

where the second equality in (2.14) follows from the consistency of $f_U(y;h)$ as an estimator for $f(y)$ (e.g., Parzen (1962)). It is clear that $\mu(\gamma)$ has a zero differential at $\gamma = \beta_U$. Now, it follows from Randles (1982, pg. 463) that $T_n(\hat{\beta})$ and $T_n(\beta_U)$ have the same limiting distribution in the design space and the proof is complete. □

The design-based properties of the estimator $\hat{f}_{dl}(y;h)$ are summarized in the following theorem.

Theorem 2.1. *Suppose Assumptions A.3 and A.4 in Section 2.2.2 hold and $K(x) \leq M$ for all x. Then, the estimator $\hat{f}_{dl}(y;h)$ is asymptotically design-unbiased (ADU) and design-consistent (in the MSE sense) for $f_U(y;h)$ which is defined in* (2.6).

Proof: Using Lemma 2.2, we have

$$\begin{aligned}
E_{\mathcal{P}}\left[\hat{f}_{dl}(y;h)\right] &\approx E_{\mathcal{P}}\left[\tilde{f}_{dl}(y;h)\right] \\
&= \frac{1}{N}E_{\mathcal{P}}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i)-K_h(y-\beta_U x_i)\right\}+\sum_{i\in U}K_h(y-\beta_U X_i)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\left[\sum_{i\in U} E_{\mathcal{P}}(I_i)d_i\left\{K_h(y-Y_i)-K_h(y-\beta_U X_i)\right\}+\sum_{i\in U}K_h(y-\beta_U X_i)\right] \\
&= \frac{1}{N}\left[\sum_{i\in U}\left\{K_h(y-Y_i)-K_h(y-\beta_U X_i)\right\}+\sum_{i\in U}K_h(y-\beta_U X_i)\right] \\
&= \frac{1}{N}\sum_{i\in U}K_h(y-Y_i) = f_U(y;h).
\end{aligned}
$$

Therefore, $\hat{f}_{dl}(y;h)$ is ADU for $f_U(y;h)$. It remains to show that the design-variance of $\hat{f}_{dl}(y;h)$ approaches zero in the limit. For this, note that using Lemma 2.2, we can write

$$
\begin{aligned}
V_{\mathcal{P}}\left[\hat{f}_{dl}(y;h)\right] &\approx V_{\mathcal{P}}\left[\tilde{f}_{dl}(y;h)\right] \\
&= \frac{1}{N^2}V_{\mathcal{P}}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i)-K_h(y-\beta_U x_i)\right\}+\sum_{i\in U}K_h(y-\beta_U X_i)\right] \\
&= \frac{1}{N^2}V_{\mathcal{P}}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i)-K_h(y-\beta_U x_i)\right\}\right] \\
&= \frac{1}{N^2}\sum\sum_{i<j\in U}(\pi_i\pi_j-\pi_{ij})\left\{\frac{K_h(y-Y_i)}{\pi_i}-\frac{K_h(y-\beta_U X_i)}{\pi_j}\right\}^2 \qquad (2.15) \\
&\leq \frac{1}{N^2}\max_{i<j\in U}|\pi_i\pi_j-\pi_{ij}|\sum\sum_{i<j\in U}\left[\left\{\frac{K_h(y-Y_i)}{\pi_i}\right\}^2+\left\{\frac{K_h(y-\beta_U X_i)}{\pi_j}\right\}^2\right] \\
&\leq \frac{1}{N^2}\max_{i<j\in U}|\pi_i\pi_j-\pi_{ij}|\sum\sum_{i<j\in U}\left[\left\{\frac{h^{-1}M}{\lambda}\right\}^2+\left\{\frac{h^{-1}M}{\lambda}\right\}^2\right] \\
&= n\max_{i<j\in U}|\pi_i\pi_j-\pi_{ij}|\frac{1}{nh^2}\frac{(N-1)}{N}\frac{M^2}{\lambda^2} \longrightarrow 0.
\end{aligned}
$$

To reach (2.15), we used the fact that the second summation in the first equality above is a finite population quantity and, hence, considered fixed with respect to the randomization distribution so its variance is zero. Define $u_i(h) = K_h(y-y_i)$ and $v_i(h) = K_h(y-\beta_U x_i)$. Then, (2.15) follows because $\sum_{i\in s} d_i\left\{u_i(h)-v_i(h)\right\}$ is the Horvitz-Thompson estimator of the finite population total of $\{u(h)-v(h)\}$. The zero limit follows from assumptions A.3 and A.4. $\square$

Next, we give the main theorem which states the bias and MISE formulae of the estimator $\hat{f}_{dl}(y;h)$ under the combined design-model-based mode of inference.

Theorem 2.2. *Suppose Assumptions A.1–A.4 hold. The bias and the MISE of* $\hat{f}_{dl}(y;h)$, *under the combined inference approach, are given by:*

$$Bias_C\left[\hat{f}_{dl}(y;h)\right] = \frac{1}{2}h^2 c_K f''(y) + o(h^2), \tag{2.16}$$

and

$$MISE_C\left[\hat{f}_{dl}(\cdot;h)\right] = \frac{\delta}{nh^3}\mu_{Y^2}\left(1 - \frac{\mu_{XY}^2}{\mu_{X^2}\mu_{Y^2}}\right)d_{K'} + \frac{1}{4}h^4 c_K^2 d_{f''} + o\left(h^4 + \frac{1}{Nh^3}\right), \tag{2.17}$$

where $\mu_{V^r} = E_\xi(V^r)$ *and* $\delta = nN^{-2}\sum_{i\in U}\Delta_i$ *with* $\Delta_i = (1-\pi_i)/\pi_i$.

Remarks: There are some interesting observations to be taken from Theorem 2.2.

(i) The leading term in the bias of the estimator $\hat{f}_{dl}(y;h)$ is identical to the leading term in the bias of the standard IID kernel density estimator that does not use any auxiliary information (see Section 1.2). This means that the auxiliary information and the sampling scheme do not affect the leading term in the bias of kernel density estimators. Bellhouse and Stafford (1999) had the same observation when they studied design-weighted kernel density estimators which use the sampling weights but do not use any auxiliary information.

(ii) The first term in the MISE of the estimator $\hat{f}_{dl}(y;h)$ has a very interesting form. First, it is readily seen that if both X and Y are standardized, the first term in the MISE reduces to $(nh^3)^{-1}\delta(1-\rho_{XY}^2)d_{K'}$ where ρ_{XY} is the correlation coefficient between X and Y. Thus, the MISE of the proposed estimator is a decreasing function of the correlation between the study variable Y and the auxiliary data X gets stronger. Second, the sampling design effect is represented in the MISE formula via the quantity $\delta = nN^{-2}\sum_{i\in U}(1-\pi_i)/\pi_i$.

(iii) Interestingly, unlike the unit power we usually get for the bandwidth in the denominator of the asymptotic variance of standard kernel density estimators that do not use any auxiliary data (see Eq. (1.5)), we have a third power for h in the asymptotic variance of $\hat{f}_{dl}(y;h)$ (see Eq. (2.17)). This result is due to the bivariate setting imposed in

$\hat{f}_{dl}(y;h)$ through the auxiliary data. Smoothing parameters of similar order were obtained by Ahmad (2002) when dealing with the problem of kernel density estimation in a continuous randomized response model which involves a similar bivariate setting.

Proof: We start with the bias statement. Using rules of conditional expectation, we have

$$E_C\left[\hat{f}_{dl}(y;h)\right] = E_\xi\left\{E_\mathcal{P}\left[\hat{f}_{dl}(y;h)|\mathbf{X}_U,\mathbf{Y}_U\right]\right\}.$$

Using Lemma 2.2, it can be shown that (see the proof of Theorem 2.1)

$$E_\mathcal{P}\left[\hat{f}_{dl}(y;h)\right] \approx E_\mathcal{P}\left[\tilde{f}_{dl}(y;h)\right] = f_U(y;h). \tag{2.18}$$

Therefore,

$$\begin{aligned} E_C\left[\hat{f}_{dl}(y;h)\right] &= \frac{1}{N}\sum_{i\in U} E_\xi\left[K_h(y-Y_i)\right] \\ &\stackrel{iid}{=} E_\xi\left[K_h(y-Y_1)\right] \\ &= \frac{1}{h}\int_\mathbb{R} K\left(\frac{y_1-y}{h}\right) f(y_1)dy_1, \end{aligned} \tag{2.19}$$

where (2.19) follows from the symmetry of the kernel $K(\cdot)$. Using a simple change of variables under the integral in (2.19), we get

$$\begin{aligned} E_C\left[\hat{f}_{dl}(y;h)\right] &= \int_\mathbb{R} K(z)f(y+hz)dz \\ &= \int_\mathbb{R} K(z)\left\{f(y)+hzf'(y)+\frac{1}{2}h^2z^2f''(y)+\dots\right\}dz \\ &= f(y)+\frac{1}{2}h^2f''(y)\int_\mathbb{R} z^2K(z)dz+o(h^2) \\ &= f(y)+\frac{1}{2}h^2c_Kf''(y)+o(h^2), \end{aligned} \tag{2.20}$$

where the last equality follows from Assumptions A.2(i–ii).

Next, we work on the variance of the estimator $\hat{f}_{dl}(y)$. Again using rules of conditional expectations, we can write

$$V_C\left[\hat{f}_{dl}(y)\right] = E_\xi\left\{V_{\mathcal{P}}\left[\hat{f}_{dl}(y)|\mathbf{X}_U,\mathbf{Y}_U\right]\right\} + V_\xi\left\{E_{\mathcal{P}}\left[\hat{f}_{dl}(y)|\mathbf{X}_U,\mathbf{Y}_U\right]\right\} := I_1 + I_2. \quad (2.21)$$

From (2.18), we have

$$\begin{aligned} I_2 &= V_\xi\left[f_U(y;h)\right] = V_\xi\left[\frac{1}{N}\sum_{i\in U} K_h(y-Y_i)\right] \\ &\overset{iid}{=} \frac{1}{N} V_\xi\left[K_h(y-Y_1)\right] = (Nh)^{-1} d_K f(y) + o\{(Nh)^{-1}\}, \end{aligned} \quad (2.22)$$

where (2.22) is a standard result in kernel density estimation (cf. Wand and Jones (1995)).

On the other hand, to evaluate the first term in (2.21), we start by finding the approximate design variance of $\hat{f}_{dl}(\cdot)$ as follows. From Lemma 2.2, we can write

$$\begin{aligned} V_{\mathcal{P}}\left[\hat{f}_{dl}(y;h)\right] &\approx V_{\mathcal{P}}\left[\tilde{f}_{dl}(y;h)\right] \\ &= \frac{1}{N^2} V_{\mathcal{P}}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i) - K_h(y-\beta_U x_i)\right\}\right] \\ &= \frac{1}{N^2} \sum\sum_{i,j\in U}\left[\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j}\left\{K_h(y-Y_i) - K_h(y-\beta_U X_i)\right\}\times\right. \\ &\qquad\qquad \left.\left\{K_h(y-Y_j) - K_h(y-\beta_U X_j)\right\}\right]. \end{aligned} \quad (2.23)$$

The equality in (2.23) follows because $\sum_{i\in s} d_i\left\{K_h(y-y_i) - K_h(y-\beta_U x_i)\right\}$ is the Horvitz-Thompson estimator of the finite population total of $\{K_h(y-y_i) - K_h(y-\beta_U x_i)\}$. Taking $\Delta_{ij} = (\pi_{ij}-\pi_i\pi_j)/\pi_i\pi_j$, $\Delta_i = (1-\pi_i)/\pi_i$ and $\hat{Y}_{Ui} = \beta_U X_i$, (2.23) can be rewritten as follows;

$$\begin{aligned} V_{\mathcal{P}}\left[\hat{f}_{dl}(y;h)\right] &\approx \frac{1}{N^2}\sum_{i\in U}\Delta_i\left\{K_h(y-Y_i) - K_h(y-\hat{Y}_{Ui})\right\}^2 \\ &\quad + \frac{1}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij}\left\{K_h(y-Y_i) - K_h(y-\hat{Y}_{Ui})\right\}\times \\ &\qquad\qquad \left\{K_h(y-Y_j) - K_h(y-\hat{Y}_{Uj})\right\} \\ &= \frac{1}{N^2}\sum_{i\in U}\Delta_i V_i^2 + \frac{1}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij} V_i V_j, \end{aligned} \quad (2.24)$$

where $V_l = \left\{K_h(y-Y_l) - K_h(y-\hat{Y}_{Ul})\right\}$. Then, from (2.21) and (2.24), we have

$$I_1 \approx \frac{1}{N^2}\sum_{i\in U}\Delta_i E_\xi(V_i^2) + \frac{1}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij}E_\xi(V_iV_j). \tag{2.25}$$

We work each term in (2.25) separately. First, note that by a Taylor expansion,

$$K_h(\hat{Y}_{Ui}-y) \approx \frac{1}{h}K\left(\frac{Y_i-y}{h}\right) + \frac{(\hat{Y}_{Ui}-Y_i)}{h^2}K'\left(\frac{Y_i-y}{h}\right),$$

and

$$\begin{aligned} K_h^2(\hat{Y}_{Ui}-y) \approx{}& \frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right) + 2\frac{(\hat{Y}_{Ui}-Y_i)}{h^3}K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_i-y}{h}\right) \\ &+\frac{(\hat{Y}_{Ui}-Y_i)^2}{h^4}\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2. \end{aligned}$$

Using these expansions, we can write

$$\begin{aligned} V_i^2 &= K_h^2(Y_i-y) - 2K_h(Y_i-y)K_h(\hat{Y}_{Ui}-y) + K_h^2(\hat{Y}_{Ui}-y) \\ &= \frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right) - 2\left[\frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right) + \frac{(\hat{Y}_{Ui}-Y_i)}{h^3}K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_i-y}{h}\right)\right] \\ &\quad+\frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right) + 2\frac{(\hat{Y}_{Ui}-Y_i)}{h^3}K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_i-y}{h}\right) \\ &\quad+\frac{(\hat{Y}_{Ui}-Y_i)^2}{h^4}\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2 \\ &= \frac{1}{h^4}\hat{Y}_{Ui}^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2 - \frac{2}{h^4}\hat{Y}_{Ui}Y_i\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2 \\ &\quad+\frac{1}{h^4}Y_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2 := W_1 - 2W_2 + W_3. \end{aligned} \tag{2.26}$$

We now evaluate the expectation of each term on the right-hand side of (2.26), in reverse.

$$\begin{aligned} \frac{1}{N}E_\xi(W_3) &= \frac{1}{Nh^4}\int_{\mathbb{R}} y_i^2\left[K'\left(\frac{y_i-y}{h}\right)\right]^2 f(y_i)dy_i \\ &= \frac{1}{Nh^3}\int_{\mathbb{R}}(y+hz)^2\left[K'(z)\right]^2 f(y+hz)dz \\ &= \frac{1}{Nh^3}\int_{\mathbb{R}}\left[K'(z)\right]^2(y^2+2yhz+h^2z^2)\left[f(y)+hzf'(y)+\ldots\right]dz \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{Nh^3} y^2 f(y) \int_{\mathbb{R}} \left[K'(z)\right]^2 dz + \frac{1}{Nh^2}\left[y^2 f'(y) + 2yf(y)\right] \int_{\mathbb{R}} z\left[K'(z)\right]^2 dz \\
&\quad + \frac{1}{Nh}\left[f(y) + 2yf'(y)\right] \int_{\mathbb{R}} z^2 \left[K'(z)\right]^2 dz + o\left(\frac{1}{Nh}\right) \\
&= \frac{1}{Nh^3} y^2 f(y) d_{K'} + \frac{1}{Nh}\left[f(y) + 2yf'(y)\right] c^*_{K'} + o\left(\frac{1}{Nh}\right).
\end{aligned} \tag{2.27}
$$

where (2.27) follows from the fact that $\int_{\mathbb{R}} z\{K'(z)\}^2 dz = 0$. For W_2, observe that

$$
\begin{aligned}
\frac{1}{N} E_\xi(W_2) &= \frac{1}{Nh^4} E_\xi \left\{ \beta_U X_i Y_i \left[K'\left(\frac{Y_i - y}{h}\right)\right]^2 \right\} \\
&= \frac{1}{Nh^4} E_\xi \left\{ \left(\sum_{j \in U} X_j Y_j \Big/ \sum_{j \in U} X_j^2 \right) X_i Y_i \left[K'\left(\frac{Y_i - y}{h}\right)\right]^2 \right\} \\
&= \frac{1}{Nh^4} E_\xi \left\{ \left(\sum_{j \in U} X_j Y_j \right) X_i Y_i \left[K'\left(\frac{Y_i - y}{h}\right)\right]^2 \Big/ \sum_{j \in U} X_j^2 \right\} \\
&:= \frac{1}{Nh^4} E_\xi \left\{ \frac{A_1}{B_1} \right\} \approx \frac{1}{Nh^4} \frac{E_\xi(A_1)}{E_\xi(B_1)}.
\end{aligned} \tag{2.28}
$$

Clearly,

$$
A_1 = X_i^2 Y_i^2 \left[K'\left(\frac{Y_i - y}{h}\right)\right]^2 + \left(\sum_{j \in U, j \neq i} X_j Y_j \right) X_i Y_i \left[K'\left(\frac{Y_i - y}{h}\right)\right]^2 := A_{11} + A_{12}.
$$

Starting with A_{11}, we have

$$
\begin{aligned}
\frac{1}{Nh^4} E_\xi(A_{11}) &= \frac{1}{Nh^4} \iint_{\mathbb{R}} x_i^2 y_i^2 \left[K'\left(\frac{y_i - y}{h}\right)\right]^2 t(x_i|y_i) f(y_i) dx_i dy_i \\
&= \frac{1}{Nh^3} \iint_{\mathbb{R}} w^2 (y + hz)^2 \left[K'(z)\right]^2 t(w|y + hz) f(y + hz) dw dz \\
&= \frac{1}{Nh^3} y^2 f(y) d_{K'} \int_{\mathbb{R}} w^2 t(w|y) dw + \frac{1}{Nh} c^*_{K'} \times \\
&\quad \int_{\mathbb{R}} w^2 \left[t(w|y) f(y) + y^2 f'(y) t_2'(w|y) + 2yf'(y) t(w|y) \right. \\
&\qquad \left. + 2yf(y) t_2'(w|y) \right] dw + o\left(\frac{1}{Nh}\right).
\end{aligned}
$$

For A_{12}, note that

$$
\frac{1}{Nh^4} E_\xi(A_{12}) = E_\xi \left(\sum_{j \in U, j \neq i} X_j Y_j \right) \frac{1}{Nh^4} \iint_{\mathbb{R}} x_i y_i \left[K'\left(\frac{y_i - y}{h}\right)\right]^2 t(x_i|y_i) f(y_i) dx_i dy_i
$$

$$
\begin{aligned}
&= (N-1)\mu_{XY}\frac{1}{Nh^3}\iint_{\mathbb{R}} w(y+hz)\left[K'(z)\right]^2 t(w|y+hz)f(y+hz)dwdz \\
&= (N-1)\mu_{XY}\left\{\frac{1}{Nh^3}yf(y)d_{K'}\int_{\mathbb{R}} wt(w|y)dw+\frac{1}{Nh}c^*_{K'}\times\right. \\
&\quad \left.\int_{\mathbb{R}} w\left[yf'(y)t_2'(w|y)+f'(y)t(w|y)+f(y)t_2'(w|y)\right]dw+o\left(\frac{1}{Nh}\right)\right\}.
\end{aligned}
$$

On the other hand,

$$
E_\xi(B_1) \overset{iid}{=} NE_\xi(X^2) = N\mu_{X^2}.
$$

Substituting these results in (2.28), we get

$$
\begin{aligned}
\frac{1}{N}E_\xi(W_2) = \frac{1}{Nh^3}\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}yf(y)d_{K'}\int_{\mathbb{R}} xt(x|y)dx+\frac{1}{Nh}\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}c^*_{K'}\times \\
\int_{\mathbb{R}} x\left[yf'(y)t_2'(x|y)+f'(y)t(x|y)+f(y)t_2'(x|y)\right]dx+o\left(\frac{1}{Nh}\right).
\end{aligned}
\tag{2.29}
$$

We now consider W_1. First, observe that

$$
\begin{aligned}
\frac{1}{N}E_\xi(W_1) &= \frac{1}{Nh^4}E_\xi\left\{\beta_U^2 X_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2\right\} \\
&= \frac{1}{Nh^4}E_\xi\left\{\left(\sum_{j\in U}X_jY_j/\sum_{j\in U}X_j^2\right)^2 X_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2\right\} \\
&= \frac{1}{Nh^4}E_\xi\left\{\left(\sum_{j\in U}X_jY_j\right)^2 X_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2\Big/\left(\sum_{j\in U}X_j^2\right)^2\right\} \\
&:= \frac{1}{Nh^4}E_\xi\left\{\frac{A_2}{B_2}\right\}\approx\frac{1}{Nh^4}\frac{E_\xi(A_2)}{E_\xi(B_2)}.
\end{aligned}
\tag{2.30}
$$

Second, expand the squared sum in A_2 to get

$$
\begin{aligned}
A_2 &= \left(\sum_{j\in U}X_j^2Y_j^2+\sum\sum_{j,k\in U,j\neq k}X_jY_jX_kY_k\right)X_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2 \\
&= X_i^4Y_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2+\left(\sum_{j\in U,j\neq i}X_j^2Y_j^2\right)X_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&+2X_i^3Y_i\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2\sum_{j\neq i}X_jY_j+\left(\sum\sum_{j,k\in U,j\neq k\neq i}X_jY_jX_kY_k\right)X_i^2\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2\\
:=&A_{21}+A_{22}+A_{23}+A_{24}.
\end{aligned}
$$

Next, we evaluate the expectation of each of the four terms, in order, as follows.

$$
\begin{aligned}
\frac{1}{Nh^4}E_\xi(A_{21}) &= \frac{1}{Nh^4}\iint_{\mathbb{R}} x_i^4y_i^2\left[K'\left(\frac{y_i-y}{h}\right)\right]^2 t(x_i|y_i)f(y_i)dx_idy_i\\
&= \frac{1}{Nh^3}\iint_{\mathbb{R}} w^4(y+hz)^2\left[K'(z)\right]^2 t(w|y+hz)f(y+hz)dwdz\\
&= \frac{1}{Nh^3}y^2f(y)d_{K'}\int_{\mathbb{R}} w^4t(w|y)dw+\frac{1}{Nh}c^*_{K'}\times\\
&\quad\int_{\mathbb{R}} w^4\left[t(w|y)f(y)+y^2f'(y)t_2'(w|y)+2yf'(y)t(w|y)\right.\\
&\qquad\qquad\left.+2yf(y)t_2'(w|y)\right]dw+o\left(\{Nh\}^{-1}\right).
\end{aligned}
$$

For A_{22}, note that

$$
\begin{aligned}
\frac{1}{Nh^4}E_\xi(A_{22}) &= \frac{1}{Nh^4}E_\xi\left(\sum_{j\in U,j\neq i}X_j^2Y_j^2\right)\iint_{\mathbb{R}} x_i^2\left[K'\left(\frac{y_i-y}{h}\right)\right]^2 t(x_i|y_i)f(y_i)dx_idy_i\\
&= (N-1)E_\xi\left(X_1^2Y_1^2\right)\frac{1}{Nh^3}\iint_{\mathbb{R}} w^2\left[K'(z)\right]^2 t(w|y+hz)f(y+hz)dwdz\\
&= (N-1)\mu_{X^2Y^2}\left\{\frac{1}{Nh^3}f(y)d_{K'}\int_{\mathbb{R}} w^2t(w|y)dw+\frac{1}{Nh}c^*_{K'}f'(y)\times\right.\\
&\qquad\qquad\left.\int_{\mathbb{R}} w^2t_2'(w|y)dw+o\left(\frac{1}{Nh}\right)\right\}.
\end{aligned}
$$

Considering A_{23}, we have

$$
\begin{aligned}
\frac{1}{Nh^4}E_\xi(A_{23}) &= \frac{2}{Nh^4}(N-1)E_\xi\left(X_1Y_1\right)\iint_{\mathbb{R}} x_i^3y_i\left[K'\left(\frac{y_i-y}{h}\right)\right]^2 t(x_i|y_i)f(y_i)dx_idy_i\\
&= \frac{2}{Nh^3}(N-1)\mu_{XY}\iint_{\mathbb{R}} w^3(y+hz)\left[K'(z)\right]^2 t(w|y+hz)f(y+hz)dwdz\\
&= 2(N-1)\mu_{XY}\left\{\frac{1}{Nh^3}yf(y)d_{K'}\int_{\mathbb{R}} w^3t(w|y)dw+\frac{1}{Nh}c^*_{K'}\times\right.\\
&\quad\left.\int_{\mathbb{R}} w^3\left[yf'(y)t_2'(w|y)+f'(y)t(w|y)+f(y)t_2'(w|y)\right]dw+o\left(\frac{1}{Nh}\right)\right\}.
\end{aligned}
$$

For A_{24}, we have

$$
\begin{aligned}
&\frac{1}{Nh^4}E_\xi(A_{24})\\
&=\frac{1}{Nh^4}E_\xi\left(\sum\sum_{j,k\in U,j\neq k\neq i}X_jY_jX_kY_k\right)\iint_{\mathbb{R}}x_i^2\left[K'\left(\frac{y_i-y}{h}\right)\right]^2t(x_i|y_i)f(y_i)dx_idy_i\\
&=\frac{[N(N-1)-2(N-1)]}{Nh^3}E_\xi^2(X_1Y_1)\iint_{\mathbb{R}}w^2\left[K'(z)\right]^2t(w|y+hz)f(y+hz)dwdz\\
&=(N-1)(N-2)\mu_{XY}^2\left\{\frac{1}{Nh^3}f(y)d_{K'}\int_{\mathbb{R}}w^2t(w|y)dw+\frac{1}{Nh}c_{K'}^*f'(y)\times\right.\\
&\qquad\qquad\left.\int_{\mathbb{R}}w^2t_2'(w|y)dw+o\left(\frac{1}{Nh}\right)\right\}.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
E_\xi(B_2)&=E_\xi\left(\sum_{j\in U}X_j^2\right)^2\\
&=E_\xi\left(\sum_{j\in U}X_j^4+\sum\sum_{j,k\in U,j\neq k}X_j^2X_k^2\right)\\
&\overset{iid}{=}NE_\xi\left(X_1^4\right)+N(N-1)E_\xi\left(X_1^2X_2^2\right)\\
&=N\mu_{X^4}+N(N-1)\mu_{X^2}^2=N(N-1)[\mu_{X^2}^2+\mu_{X^4}/(N-1)].
\end{aligned}
$$

Substituting these results into (2.30), we get

$$
\begin{aligned}
\frac{1}{N}E_\xi(W_1)=&\frac{1}{Nh^3}\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{[\mu_{X^2}^2+\mu_{X^4}/(N-1)]}d_{K'}f(y)\int_{\mathbb{R}}w^2t(w|y)dw\\
&+\frac{1}{Nh}\left(\frac{N-2}{N}\right)\frac{c_{K'}^*}{[\mu_{X^2}^2+\mu_{X^4}/(N-1)]}f'(y)\int_{\mathbb{R}}w^2t_2'(w|y)dw+o\left(\frac{1}{Nh}\right).
\end{aligned}
$$

Since $\mu_{X^2}^2+\{\mu_{X^4}/(N-1)\}=\mu_{X^2}^2+o(1)$ and we are mainly interested in asymptotic results, we can write

$$
\begin{aligned}
\frac{1}{N}E_\xi(W_1)=&\frac{1}{Nh^3}\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}d_{K'}f(y)\int_{\mathbb{R}}w^2t(w|y)dw\\
&+\frac{1}{Nh}\left(\frac{N-2}{N}\right)\frac{c_{K'}^*}{\mu_{X^2}^2}f'(y)\int_{\mathbb{R}}w^2t_2'(w|y)dw+o\left(\frac{1}{Nh}\right).
\end{aligned}
\tag{2.31}
$$

Now, we use (2.27), (2.29) and (2.31) in (2.26) to get

$$\begin{aligned}\frac{1}{N}E_{\xi}(V_i^2) = \frac{1}{Nh^3}\Bigg[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}f(y)\int_{\mathbb{R}}x^2t(x|y)dx - 2\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}yf(y)\times \\ \int_{\mathbb{R}}xt(x|y)dx + y^2f(y)\Bigg]d_{K'} + \frac{1}{Nh}\Bigg[\left(\frac{N-2}{N}\right)\frac{1}{\mu_{X^2}^2}f'(y)\times \\ \int_{\mathbb{R}}x^2t_2'(x|y)dx - 2\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}\int_{\mathbb{R}}w\left[yf'(y)t_2'(x|y)+f'(y)\times\right. \\ \left. t(x|y)+f(y)t_2'(x|y)\right]dx + f(y) + 2yf'(y)\Bigg]c_{K'}^* + o\left(\frac{1}{Nh}\right). \qquad (2.32)\end{aligned}$$

If we keep only terms of order Nh^3, (2.32) reduces to

$$\begin{aligned}\frac{1}{N}E_{\xi}(V_i^2) = \frac{1}{Nh^3}\Bigg[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}f(y)\int_{\mathbb{R}}x^2t(x|y)dx - 2\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}yf(y)\times \\ \int_{\mathbb{R}}xt(x|y)dx + y^2f(y)\Bigg]d_{K'} + o\left(\frac{1}{Nh^3}\right). \qquad (2.33)\end{aligned}$$

To evaluate the second term in (2.25), note that by a Taylor expansion for $K_h(y-\hat{Y}_{Ul})$ about Y_l, we have

$$\begin{aligned}V_iV_j &= \left\{K_h(y-Y_i)-K_h(y-\hat{Y}_{Ui})\right\}\left\{K_h(y-Y_j)-K_h(y-\hat{Y}_{Uj})\right\} \\ &= K_h(y-Y_i)K_h(y-Y_j) - 2K_h(y-Y_i)K_h(y-\hat{Y}_{Uj}) + K_h(y-\hat{Y}_{Ui})K_h(y-\hat{Y}_{Uj}) \\ &\approx \frac{1}{h^2}K\left(\frac{Y_i-y}{h}\right)K\left(\frac{Y_j-y}{h}\right) - \frac{2}{h^2}K\left(\frac{Y_i-y}{h}\right)K\left(\frac{Y_j-y}{h}\right) - \frac{2}{h^3}(\hat{Y}_{Uj}-Y_j)\times \\ &\quad K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) + \frac{1}{h^2}K\left(\frac{Y_i-y}{h}\right)K\left(\frac{Y_j-y}{h}\right) + \frac{2}{h^3}(\hat{Y}_{Uj}-Y_j)\times \\ &\quad K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) + \frac{1}{h^4}(\hat{Y}_{Ui}-Y_i)(\hat{Y}_{Uj}-Y_j)K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\ &= \frac{1}{h^4}\left(\hat{Y}_{Ui}\hat{Y}_{Uj}-2Y_i\hat{Y}_{Uj}+Y_iY_j\right)K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\ &= \frac{1}{h^4}\hat{Y}_{Ui}\hat{Y}_{Uj}K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) - 2\frac{1}{h^4}Y_i\hat{Y}_{Uj}K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\ &\quad + \frac{1}{h^4}Y_iY_jK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\ &:= H_1 - 2H_2 + H_3. \qquad (2.34)\end{aligned}$$

We now examine the expectation of each of the three terms in reverse. For H_3, we have

$$
\begin{aligned}
\frac{1}{n}E_\xi(H_3) &= \frac{1}{nh^4}E_\xi\left[Y_iY_jK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\
&\overset{iid}{=} \frac{1}{nh^4}\left\{E_\xi\left[Y_1K'\left(\frac{Y_1-y}{h}\right)\right]\right\}^2 \\
&= \frac{1}{nh^4}\left\{\int_{\mathbb{R}} y_1K'\left(\frac{y_1-y}{h}\right)f(y_1)dy_1\right\}^2 \\
&= \frac{1}{nh^4}\left\{h\int_{\mathbb{R}}(y+hz)K'(z)f(y+hz)dz\right\}^2 \\
&= \frac{1}{nh^2}\left\{yf(y)\int_{\mathbb{R}}K'(z)dz+h[f(y)+yf'(y)]\int_{\mathbb{R}}zK'(z)dz+\cdots\right\}^2 \\
&= \frac{1}{n}\left\{-[f(y)+yf'(y)]+\cdots\right\}^2 = O\left(\frac{1}{n}\right), \qquad (2.35)
\end{aligned}
$$

where (2.35) follows from the facts that $\int_{\mathbb{R}}K'(z)dz = 0$ and $\int_{\mathbb{R}}zK'(z)dz = -1$ by the assumptions on K. Next, we consider H_2.

$$
\begin{aligned}
\frac{1}{n}E_\xi(H_2) &= \frac{1}{nh^4}E_\xi\left[\beta_UX_jY_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\
&= \frac{1}{nh^4}E_\xi\left[\left(\sum_{k\in U}X_kY_k\Big/\sum_{k\in U}X_k^2\right)X_jY_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\
&= \frac{1}{nh^4}E_\xi\left[\left(\sum_{k\in U}X_kY_K\right)X_jY_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\Big/\sum_{k\in U}X_k^2\right] \\
&:= \frac{1}{nh^4}E_\xi\left[\frac{A_3}{B_3}\right] \approx \frac{1}{nh^4}\frac{E_\xi(A_3)}{E_\xi(B_3)}. \qquad (2.36)
\end{aligned}
$$

First, observe that

$$
\begin{aligned}
A_3 = X_j^2Y_jY_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) + X_jX_iY_i^2K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\
+\left(\sum_{k\in U,k\neq i\neq j}X_kY_k\right)X_jY_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) := A_{31}+A_{32}+A_{33}.
\end{aligned}
$$

Second, we evaluate the expectation of the three terms in order. Considering A_{31}, we have

$$
\frac{1}{nh^4}E_\xi(A_{31})
$$

$$
\begin{aligned}
&\overset{iid}{=} \frac{1}{nh^4} E_\xi \left[Y_1 K' \left(\frac{Y_1 - y}{h} \right) \right] E_\xi \left[X_2^2 Y_2 K' \left(\frac{Y_2 - y}{h} \right) \right] \\
&= \frac{1}{nh^4} \left[\int_{\mathbb{R}} y_1 K' \left(\frac{y_1 - y}{h} \right) f(y_1) dy_1 \right] \left[\iint_{\mathbb{R}} x_2^2 y_2 K' \left(\frac{y_2 - y}{h} \right) t(x_2|y_2) f(y_2) dx_2 dy_2 \right] \\
&= \frac{1}{nh^2} \left[\int_{\mathbb{R}} (y + hz_1) K'(z_1) f(y + hz_1) dz_1 \right] \\
&\qquad \left[\iint_{\mathbb{R}} w^2 (y + hz_2) K'(z_2) t(w|y + hz_2) f(y + hz_2) dw dz_2 \right] \\
&= \frac{1}{nh^2} \left[h \left\{ f(y) + y f'(y) \right\} \int_{\mathbb{R}} z K'(z) dz \right] \left[h \int_{\mathbb{R}} z K'(z) dz \right. \\
&\qquad \left. \int_{\mathbb{R}} w^2 \left\{ y t_2'(w|y) f(y) + t(w|y) f(y) + y t(w|y) f'(y) \right\} dw \right] \\
&= \frac{1}{n} \left\{ f(y) + y f'(y) \right\} \int_{\mathbb{R}} w^2 \left\{ y t_2'(w|y) f(y) + t(w|y) f(y) + y t(w|y) f'(y) \right\} dw = O\left(\frac{1}{n} \right).
\end{aligned}
$$

For A_{32}, we have

$$
\begin{aligned}
&\frac{1}{nh^4} E_\xi (A_{32}) \\
&\overset{iid}{=} \frac{1}{nh^4} E_\xi \left[X_1 Y_1^2 K' \left(\frac{Y_1 - y}{h} \right) \right] E_\xi \left[X_2 K' \left(\frac{Y_2 - y}{h} \right) \right] \\
&= \frac{1}{nh^4} \left[\iint_{\mathbb{R}} x_1 y_1^2 K' \left(\frac{y_1 - y}{h} \right) t(x_1, y_1) dx_1 dy_1 \right] \left[\iint_{\mathbb{R}} x_2 K' \left(\frac{y_2 - y}{h} \right) t(x_2, y_2) dx_2 dy_2 \right] \\
&= \frac{1}{nh^2} \left[\iint_{\mathbb{R}} w_1 (y + hz_1)^2 K'(z_1) t(w_1, y + hz_1) dw_1 dz_1 \right] \\
&\left[\iint_{\mathbb{R}} w_2 K'(z_2) t(w_2, y + hz_2) dw_2 dz_2 \right] \\
&= \frac{1}{nh^2} \left[\iint_{\mathbb{R}} w_1 (y + hz_1)^2 K'(z_1) \left\{ t(y|w_1) + hz_1 t_1'(y|w_1) \right\} g(w_1) dw_1 dz_1 \right] \\
&\left[\iint_{\mathbb{R}} w_2 K'(z_2) \left\{ t(y|w_2) + hz_2 t_1'(y|w_2) \right\} g(w_2) dw_2 dz_2 \right] \\
&\approx \frac{1}{nh^2} \left[\iint_{\mathbb{R}} w_1 (y + hz_1)^2 K'(z_1) \left\{ t(y|w_1) + hz_1 t_1'(y|w_1) \right\} g(w_1) dw_1 dz_1 \right] \\
&\left[\iint_{\mathbb{R}} w_2 K'(z_2) \left\{ t(y|w_2) + hz_2 t_1'(y|w_2) \right\} g(w_2) dw_2 dz_2 \right] \\
&= \frac{1}{nh^2} \left[\iint_{\mathbb{R}} w_1 K'(z_1) \left\{ y^2 t(y|w_1) + hz_1 y^2 t_1'(y|w_1) + 2yhz_1 t(y|w_1) + 2yh^2 z_1^2 t_1'(y|w_1) \right. \right. \\
&\qquad \left. \left. + h^2 z_1^2 t(y|w_1) + h^3 z_1^3 t_1'(y|w_1) \right\} g(w_1) dw_1 dz_1 \right] \\
&\left[h \int_{\mathbb{R}} w_2 t_1'(y|w_2) g(w_2) dw_2 \int_{\mathbb{R}} z_2 K'(z_2) dz_2 \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{nh^2}\left[h\int_{\mathbb{R}} w_1\left\{y^2 t_1'(y|w_1)+2yt(y|w_1)\right\}g(w_1)dw_1\int_{\mathbb{R}} z_1K'(z_1)dz_1\right.\\
&\qquad\qquad \left.+O(h^3)\right]\left[h\int_{\mathbb{R}} w_2t_1'(y|w_2)g(w_2)dw_2\int_{\mathbb{R}} z_2K'(z_2)dz_2\right]\\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

Consider A_{33} and observe that

$$
\begin{aligned}
&\frac{1}{nh^4}E_\xi(A_{33})\\
&\overset{iid}{=} \frac{1}{nh^4}(N-2)E_\xi(X_3Y_3)E_\xi\left[Y_1K'\left(\frac{Y_1-y}{h}\right)\right]E_\xi\left[X_2K'\left(\frac{Y_2-y}{h}\right)\right]\\
&= \frac{(N-2)}{nh^4}\mu_{XY}\left[\int_{\mathbb{R}} y_1K'\left(\frac{y_1-y}{h}\right)f(y_1)dy_1\right]\left[\iint_{\mathbb{R}} x_2K'\left(\frac{y_2-y}{h}\right)t(x_2,y_2)dx_2dy_2\right]\\
&= \frac{(N-2)}{nh^2}\mu_{XY}\left[\int_{\mathbb{R}}(y+hz_1)K'(z_1)f(y+hz_1)dz_1\right]\left[\iint_{\mathbb{R}} wK'(z_2)t(w,y+hz_2)dwdz_2\right]\\
&\approx \frac{(N-2)}{nh^2}\mu_{XY}\left[h\left\{f(y)+yf'(y)\right\}\int_{\mathbb{R}} z_1K'(z_1)\,dz_1\right]\times\\
&\qquad\qquad\qquad \left[h\int_{\mathbb{R}} wt_1'(y|w)g(w)dw\int_{\mathbb{R}} z_2K'(z_2)\,dz_2\right]\\
&= \frac{(N-2)}{n}\mu_{XY}\left\{f(y)+yf'(y)\right\}\int_{\mathbb{R}} xt_1'(y|x)g(x)dx.
\end{aligned}
$$

For the denominator of (6.8), note that

$$
E_\xi(B_3)\overset{iid}{=} NE_\xi(X_1^2)=N\mu_{X^2}.
$$

Substituting all these results into (6.8), we get

$$
\begin{aligned}
\frac{1}{nh^4}E_\xi(H_2) &= \left(\frac{N-2}{N}\right)\frac{1}{n}\frac{\mu_{XY}}{\mu_{X^2}}\left\{f(y)+yf'(y)\right\}\int_{\mathbb{R}} xt_1'(y|x)g(x)dx+O\left(\frac{1}{nN}\right)\\
&= O\left(\frac{1}{n}\right). \qquad (2.37)
\end{aligned}
$$

Next, we work on H_1. First, observe that

$$
\frac{1}{nh^4}E_\xi(H_1) = \frac{1}{nh^4}E_\xi\left[\beta_U^2X_iX_jK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right]
$$

$$
\begin{aligned}
&= \frac{1}{nh^4} E_\xi \left[\left(\sum_{k\in U} X_k Y_k / \sum_{k\in U} X_k^2 \right)^2 X_i X_j K'\left(\frac{Y_i - y}{h}\right) K'\left(\frac{Y_j - y}{h}\right) \right] \\
&= \frac{1}{nh^4} E_\xi \left[\left(\sum_{k\in U} X_k Y_k \right)^2 X_i X_j K'\left(\frac{Y_i - y}{h}\right) K'\left(\frac{Y_j - y}{h}\right) \Big/ \left(\sum_{k\in U} X_k^2 \right)^2 \right] \\
&:= \frac{1}{nh^4} E_\xi \left[\frac{A_4}{B_4} \right] \approx \frac{1}{nh^4} \frac{E_\xi(A_4)}{E_\xi(B_4)}. \qquad (2.38)
\end{aligned}
$$

But,

$$
\begin{aligned}
A_4 &= \left(\sum_{k\in U} X_k^2 Y_k^2 + \sum\sum_{k,l\in U,\ k\neq l} X_k Y_k X_l Y_l \right) X_i X_j K'\left(\frac{Y_i - y}{h}\right) K'\left(\frac{Y_j - y}{h}\right) \\
&= \left[2X_i^3 Y_i^2 X_j + \left(\sum_{k\in U,\ k\neq i\neq j} X_k^2 Y_k^2 \right) X_i X_j + 2X_i^2 Y_i X_j^2 Y_j + 4 \left(\sum_{k\in U,\ k\neq i\neq j} X_k Y_k \right) \right. \\
&\quad \left. \times X_i^2 Y_i X_j + \left(\sum\sum_{k,l\in U,\ k\neq l\neq i\neq j} X_k Y_k X_l Y_l \right) X_i X_j \right] K'\left(\frac{Y_i - y}{h}\right) K'\left(\frac{Y_j - y}{h}\right) \\
&:= A_{41} + A_{42} + A_{43} + A_{44} + A_{45}.
\end{aligned}
$$

We now evaluate the expectation of each of the five terms, in order, separately. Observe that,

$$
\begin{aligned}
\frac{1}{nh^4} E_\xi(A_{41}) &= \frac{2}{nh^4} \iiiint_{\mathbb{R}} x_i^3 y_i^2 x_j K'\left(\frac{y_i - y}{h}\right) K'\left(\frac{y_j - y}{h}\right) t(x_i, y_i) t(x_j, y_j) \\
&\qquad dx_i dx_j dy_i dy_j \\
&= \frac{2}{nh^2} \iiiint_{\mathbb{R}} w_1^3 w_2 (y + hz_1)^2 K'(z_1) K'(z_2) t(w_1, y + hz_1) t(w_2, y + hz_2) \\
&\qquad dw_1 dw_2 dz_1 dz_2 \\
&\approx \frac{2}{nh^2} \iiiint_{\mathbb{R}} w_1^3 w_2 (y + hz_1)^2 K'(z_1) K'(z_2) g(w_1) g(w_2) \times \\
&\qquad \left\{ t(y|w_1) + hz_1 t_1'(y|w_1) \right\} \left\{ t(y|w_2) + hz_2 t_1'(y|w_2) \right\} dw_1 dw_2 dz_1 dz_2 \\
&= O\left(\frac{1}{n}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{nh^4} E_\xi(A_{42}) &\overset{iid}{=} \frac{1}{nh^4} \left[(N-2) E_\xi(X_3^2 Y_3^2) \right] \iiiint_{\mathbb{R}} x_1 x_2 K'\left(\frac{y_1 - y}{h}\right) K'\left(\frac{y_2 - y}{h}\right) \\
&\qquad t(x_1, y_1) t(x_2, y_2) dx_1 dx_2 dy_1 dy_2
\end{aligned}
$$

$$
\begin{aligned}
&= (N-2)\mu_{X^2Y^2}\frac{1}{nh^2}\iiiint_{\mathbb{R}} w_1w_2K'(z_1)\,K'(z_2)\,t(w_2,y+hz_2) \\
&\qquad\qquad t(w_1,y+hz_1)dw_1dw_2dz_1dz_2 \\
&\approx (N-2)\mu_{X^2Y^2}\frac{1}{nh^2}\iiiint_{\mathbb{R}} w_1w_2K'(z_1)\,K'(z_2)\,g(w_1)g(w_2)\times \\
&\quad \left\{t(y|w_1)+hz_1t_1'(y|w_1)\right\}\left\{t(y|w_2)+hz_2t_1'(y|w_2)\right\}dw_1dw_2dz_1dz_2 \\
&= (N-2)\mu_{X^2Y^2}\frac{1}{nh^2}\left[h^2\int_{\mathbb{R}} w_1g(w_1)t_1'(y|w_1)dw_1\int_{\mathbb{R}} z_1K'(z_1)\,dz_1\times\right. \\
&\qquad\left.\int_{\mathbb{R}} w_2g(w_2)t_1'(y|w_2)dw_2\int_{\mathbb{R}} z_2K'(z_2)\,dz_2\right] \\
&= (N-2)\mu_{X^2Y^2}\frac{1}{n}\left\{\int_{\mathbb{R}} w_1g(w_1)t_1'(y|w_1)dw_1\right\}^2,
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{nh^4}E_\xi(A_{43}) &\overset{iid}{=} \frac{2}{nh^4}\iiiint_{\mathbb{R}} x_1^2y_1x_2^2y_2K'\left(\frac{y_1-y}{h}\right)K'\left(\frac{y_2-y}{h}\right)t(x_1,y_1)\times \\
&\qquad\qquad t(x_2,y_2)dx_1dx_2dy_1dy_2 \\
&= \frac{2}{nh^2}\iiiint_{\mathbb{R}} w_1^2(y+hz_1)w_2^2(y+hz_2)K'(z_1)\,K'(z_2)\,t(w_1,y+hz_1) \\
&\qquad\qquad t(w_2,y+hz_2)dw_1dw_2dz_1dz_2 \\
&\approx \frac{2}{nh^2}\iiiint_{\mathbb{R}} w_1^2(y+hz_1)w_2^2(y+hz_2)K'(z_1)\,K'(z_2)\,g(w_1)g(w_2)\times \\
&\quad \left\{t(y|w_1)+hz_1t_1'(y|w_1)\right\}\left\{t(y|w_2)+hz_2t_1'(y|w_2)\right\}dw_1dw_2dz_1dz_2 \\
&= O\left(\frac{1}{n}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{nh^4}E_\xi(A_{44}) &\overset{iid}{=} \frac{4}{nh^4}\left[(N-2)E_\xi(X_3Y_3)\right]\iiiint_{\mathbb{R}} x_1^2y_1x_2K'\left(\frac{y_1-y}{h}\right)K'\left(\frac{y_2-y}{h}\right) \\
&\qquad\qquad t(x_1,y_1)t(x_2,y_2)dx_1dx_2dy_1dy_2 \\
&= 4(N-2)\mu_{XY}\frac{1}{nh^2}\iiiint_{\mathbb{R}} w_1^2(y+hz_1)w_2K'(z_1)\,K'(z_2)\,t(w_2,y+hz_2) \\
&\qquad\qquad t(w_1,y+hz_1)dw_1dw_2dz_1dz_2 \\
&\approx 4(N-2)\mu_{XY}\frac{1}{nh^2}\iiiint_{\mathbb{R}} w_1^2(y+hz_1)w_2K'(z_1)\,K'(z_2)\,g(w_1)g(w_2)\times \\
&\quad \left\{t(y|w_1)+hz_1t_1'(y|w_1)\right\}\left\{t(y|w_2)+hz_2t_1'(y|w_2)\right\}dw_1dw_2dz_1dz_2 \\
&= 4(N-2)\mu_{XY}\frac{1}{nh^2}\left[h^2\int_{\mathbb{R}} w_1^2g(w_1)\left\{t(y|w_1)+yt_1'(y|w_1)\right\}dw_1\times\right.
\end{aligned}
$$

$$\int_{\mathbb{R}} z_1 K'(z_1)\, dz_1 \int_{\mathbb{R}} w_2 g(w_2) t_1'(y|w_2) dw_2 \int_{\mathbb{R}} z_2 K'(z_2)\, dz_2 \Big]$$

$$= \frac{4}{n}(N-2)\mu_{XY} \left[\int_{\mathbb{R}} w_1^2 g(w_1) \left\{ t(y|w_1) + y t_1'(y|w_1) \right\} dw_1 \times \right.$$

$$\left. \int_{\mathbb{R}} w_2 g(w_2) t_1'(y|w_2) dw_2 \right]$$

and

$$\begin{aligned}
\frac{1}{nh^4} E_\xi(A_{45}) &\overset{iid}{=} \frac{1}{nh^4} \left[\left\{ N(N-1) - 4(N-2) - 2 \right\} E_\xi(X_3 Y_3 X_4 Y_4) \right] \times \\
&\quad \iiiint_{\mathbb{R}} x_1 x_2 K'\left(\frac{y_1 - y}{h}\right) K'\left(\frac{y_2 - y}{h}\right) t(x_1, y_1) t(x_2, y_2) dx_1 dx_2 dy_1 dy_2 \\
&= (N-2)(N-3)\mu_{XY}^2 \frac{1}{nh^2} \iiiint_{\mathbb{R}} w_1 w_2 K'(z_1) K'(z_2) t(w_1, y + hz_1) \\
&\qquad\qquad t(w_2, y + hz_2) dw_1 dw_2 dz_1 dz_2 \\
&\approx (N-2)(N-3)\mu_{XY}^2 \frac{1}{nh^2} \iiiint_{\mathbb{R}} w_1 w_2 K'(z_1) K'(z_2) g(w_1) g(w_2) \times \\
&\quad \left\{ t(y|w_1) + hz_1 t_1'(y|w_1) \right\} \left\{ t(y|w_2) + hz_2 t_1'(y|w_2) \right\} dw_1 dw_2 dz_1 dz_2 \\
&= (N-2)(N-3)\mu_{XY}^2 \frac{1}{nh^2} \left[h^2 \int_{\mathbb{R}} w_1 g(w_1) t_1'(y|w_1) dw_1 \times \right. \\
&\quad \left. \int_{\mathbb{R}} z_1 K'(z_1)\, dz_1 \int_{\mathbb{R}} w_2 g(w_2) t_1'(y|w_2) dw_2 \int_{\mathbb{R}} z_2 K'(z_2)\, dz_2 \right] \\
&= (N-2)(N-3)\mu_{XY}^2 \frac{1}{n} \left\{ \int_{\mathbb{R}} w_1 g(w_1) t_1'(y|w_1) dw_1 \right\}^2.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
E_\xi(B_4) &= E_\xi \left(\sum_{k \in U} X_k^4 + \sum_{k \in U,} \sum_{k \neq l} X_k^2 X_l^2 \right) \\
&= \sum_{k \in U} E_\xi(X_k^4) + \sum_{k \in U,} \sum_{k \neq l} E_\xi(X_k^2 X_l^2) \\
&\overset{iid}{=} N\mu_{X^4} + N(N-1)\mu_{X^2}^2 = N(N-1) \left[\mu_{X^2}^2 + (N-1)^{-1} \mu_{X^4} \right].
\end{aligned}$$

Substituting these results in (6.13), we get

$$\frac{1}{nh^4} E_\xi(H_1) = \frac{(N-2)(N-3)}{N(N-1)} \frac{\mu_{XY}^2}{\left[\mu_{X^2}^2 + (N-1)^{-1} \mu_{X^4} \right]} \frac{1}{n} \left[\int_{\mathbb{R}} x g(x) t_1'(y|x) dx \right]^2$$

$$= O\left(\frac{1}{n}\right). \tag{2.39}$$

Using (2.35), (2.37) and (2.39) in (2.34), we get

$$\frac{1}{n}E_{\xi}(V_iV_j) = O\left(\frac{1}{n}\right). \tag{2.40}$$

Note that Assumption A.4 implies that the following quantities are bounded:

$$\frac{n}{N^2}\sum_{i\in U}\Delta_i = \frac{n}{N^2}\sum_{i\in U}(\frac{1}{\pi_i}-1) \leq \frac{n}{N^2}\sum_{i\in U}(\frac{1}{\lambda}-1) = \frac{n}{N}\frac{(1-\lambda)}{\lambda} = O(1)$$

and

$$\frac{n}{N^2}\sum\sum_{i,j\in U,\, i\neq j}\Delta_{ij} = \frac{n}{N^2}\sum\sum_{i,j\in U,\, i\neq j}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} \leq \frac{1}{N^2}\sum\sum_{i,j\in U,\, i\neq j}\frac{n\, \max_{i\neq j}|\pi_{ij}-\pi_i\pi_j|}{\lambda^2} = O(1).$$

Finally, use (2.33) and (2.40) in (2.25) to get

$$\begin{aligned}
I_1 = &\left(\frac{1}{N}\sum_{i\in U}\Delta_i\right)\left[\frac{1}{Nh^3}\left\{\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}f(y)\int_{\mathbb{R}}x^2t(x|y)dx - 2\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}yf(y)\right.\right.\\
&\left.\left.\times\int_{\mathbb{R}}xt(x|y)dx + y^2f(y)\right\}d_{K'} + o\left(\frac{1}{Nh^3}\right)\right] + \left(\frac{n}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij}\right)O\left(\frac{1}{n}\right)\\
= &\left(\frac{n}{N^2}\sum_{i\in U}\Delta_i\right)\frac{1}{nh^3}\left[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}f(y)\int_{\mathbb{R}}x^2t(x|y)dx - 2\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}yf(y)\right.\\
&\left.\times\int_{\mathbb{R}}xt(x|y)dx + y^2f(y)\right]d_{K'} + o\left(\frac{1}{Nh^3}\right).
\end{aligned} \tag{2.41}$$

Now, use (2.22) and (2.25) in (2.21) to get

$$\begin{aligned}
V_C\left[\hat{f}_{dl}(y)\right] = &\left(\frac{n}{N^2}\sum_{i\in U}\Delta_i\right)\frac{1}{nh^3}\left[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}f(y)\int_{\mathbb{R}}x^2t(x|y)dx - 2\left(\frac{N-1}{N}\right)\times\right.\\
&\left.\frac{\mu_{XY}}{\mu_{X^2}}yf(y)\int_{\mathbb{R}}xt(x|y)dx + y^2f(y)\right]d_{K'} + o\left(\frac{1}{Nh^3}\right).
\end{aligned} \tag{2.42}$$

Integrating (2.42) over y gives the integrated variance of the estimator $\hat{f}_{dl}(y)$:

$$
\begin{aligned}
IV_C\left[\hat{f}_{dl}(\cdot)\right] &= \left(\frac{n}{N^2}\sum_{i\in U}\Delta_i\right)\frac{1}{nh^3}\left[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}\iint_{\mathbb{R}} x^2 t(x|y)f(y)dxdy - 2\left(\frac{N-1}{N}\right)\right.\\
&\quad \left.\times\frac{\mu_{XY}}{\mu_{X^2}}\iint_{\mathbb{R}} yxt(x|y)f(y)dxdy + \int_{\mathbb{R}} y^2 f(y)dy\right] d_{K'} + o\left(\frac{1}{Nh^3}\right)\\
&= \left(\frac{n}{N^2}\sum_{i\in U}\Delta_i\right)\frac{1}{nh^3}\left[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}\mu_{X^2} - 2\left(\frac{N-1}{N}\right)\frac{\mu_{XY}}{\mu_{X^2}}\mu_{XY} + \mu_{Y^2}\right]\\
&\quad \times d_{K'} + o\left(\frac{1}{Nh^3}\right)\\
&= \left(\frac{n}{N^2}\sum_{i\in U}\Delta_i\right)\frac{1}{nh^3}\left[\mu_{Y^2} - \frac{\mu_{XY}^2}{\mu_{X^2}}\right] d_{K'} + o\left(\frac{1}{Nh^3}\right). \qquad (2.43)
\end{aligned}
$$

Integrating the squared bias, see (2.20), gives the second part of the MISE of $\hat{f}_{dl}(y)$:

$$
\begin{aligned}
ISB_C\left[\hat{f}_{dl}(\cdot)\right] &= \frac{1}{4}h^4 c_K \int_{\mathbb{R}} \left\{f''(y)\right\}^2 dy + o(h^4)\\
&= \frac{1}{4}h^4 c_K d_{f''} + o(h^4). \qquad (2.44)
\end{aligned}
$$

Adding (2.43) to (2.44) gives the result and the proof is complete. □

Corollary 2.1. *Under the simple random sampling without replacement design, the MISE of the estimator $\hat{f}_{dl}(y)$ is given by:*

$$
\begin{aligned}
MISE_C\left[\hat{f}_{dl}(\cdot)\right] &= \frac{1}{nh^3}\left(1-\frac{n}{N}\right)\mu_{Y^2}\left(1-\frac{\mu_{XY}^2}{\mu_{X^2}\mu_{Y^2}}\right)d_{K'} + \frac{1}{4}h^4 c_K^2 d_{f''}\\
&\quad + o\left\{h^4 + (Nh^3)^{-1}\right\}. \qquad (2.45)
\end{aligned}
$$

Proof: The corollary follows immediately from Theorem 2.2 and the fact that under the simple random sampling design,

$$
\delta = \frac{n}{N^2}\sum_{i\in U}\frac{(1-\pi_i)}{\pi_i} = \frac{n}{N^2}\frac{N(1-n/N)}{n/N} = \left(1-\frac{n}{N}\right).
$$

2.4. Asymptotic Distribution of $\hat{f}_{dl}(y)$

To best of our knowledge, there does not exist a unified central limit theorem under the framework for design-based inference. The asymptotic normality of sample-based estimators, specially Horvitz-Thompson estimators, of the finite population mean has been studied for many well-known designs on a design-by-design basis (see Hájek (1960), Sen (1988), Thompson (1997) and Fuller (2009), among many others). In this section, we derive the asymptotic distribution of the proposed estimator $\hat{f}_{dl}(y)$ under both the design-based and the combined design-model-based inference frameworks when the sampling design $\mathcal{P}(\cdot)$ is simple random sampling without replacement (SRSWOR). Somewhat similar analyses requiring different design assumptions would lead to the asymptotic distribution of $\hat{f}_{dl}(y)$ under other sampling designs. Similarly, our analysis will be restricted to the SRSWOR design when we study the asymptotic distributions of the other two density estimators we propose in Chapters 3 and 4.

The following lemma gives the asymptotic distribution of a standardized version of the estimator $\hat{f}_{dl}(y;h)$, in the design space, under SRSWOR. This result can be used to make inference about the finite population smooth $f_U(y;h)$ based on $\hat{f}_{dl}(y;h)$.

Lemma 2.3. *Suppose Assumptions A.1–A.4 in Section 2.2.2 hold. Further, suppose that $K(x) \leq M$ for all $x \in \mathbb{R}$ and $\sqrt{n}h \to \infty$ as $n \to \infty$. Then, under SRSWOR, we have*

$$\frac{\hat{f}_{dl}(y;h) - f_U(y;h)}{\hat{\Gamma}_{\mathcal{P}}^{1/2}} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1) \tag{2.46}$$

where $\mathscr{L}_{\mathcal{P}}$ means "convergence in law in the design space $(S, \mathscr{S}, P_{\mathcal{P}})$" and

$$\hat{\Gamma}_{\mathcal{P}} = \left(1 - \frac{n}{N}\right) \frac{\left[\sum_{i\in s}(u_i(h) - z_i(h))^2 - n^{-1}\{\sum_{i\in s}(u_i(h) - z_i(h))\}^2\right]}{n(n-1)},$$

with $u_i(h) = K_h(y - y_i)$ and $z_i(h) = K_h(y - \hat{\beta}x_i)$.

Proof: First, recall the pseudo density estimator in (2.13)

$$\begin{aligned}\tilde{f}_{dl}(y;h) &= \frac{1}{N}\left[\sum_{i\in s} d_i\{K_h(y-y_i) - K_h(y-\beta_U x_i)\} + \sum_{i\in U} K_h(y-\beta_U x_i)\right] \\ &= \frac{1}{N}\left[\sum_{i\in s} d_i\{u_i(h) - v_i(h)\} + \sum_{i\in U} v_i(h)\right],\end{aligned}$$

where $v_i(h) = K_h(y-\beta_U x_i)$. Under SRSWOR, $d_i = N/n$ for all $i \in U$ and, hence,

$$\tilde{f}_{dl}(y;h) = \frac{1}{n}\sum_{i\in s}\{u_i(h) - v_i(h)\} + \frac{1}{N}\sum_{i\in U} v_i(h) = \frac{1}{n}\sum_{i\in s} w_i(h), \tag{2.47}$$

where $w_i(h) = u_i(h) - v_i(h) + (1/N)\sum_{j\in U} v_j(h)$. The design-variance of $\tilde{f}_{dl}(y;h)$ is

$$\Gamma_{\mathcal{P}} = \left(1 - \frac{n}{N}\right)\frac{\left[\sum_{i\in U}(u_i(h) - v_i(h))^2 - N^{-1}\{\sum_{i\in U}(u_i(h) - v_i(h))\}^2\right]}{n(N-1)}. \tag{2.48}$$

Since $\tilde{f}_{dl}(y;h)$ is a sample mean, as shown in Eq. (2.47), to show that

$$\frac{\tilde{f}_{dl}(y;h) - f_U(y;h)}{\Gamma_{\mathcal{P}}^{1/2}} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1), \tag{2.49}$$

we need to verify the following Lyapunov's condition (e.g., Thompson (1997, pg. 59)):

$$\left(1 - \frac{n}{N}\right)\sum_{i\in s} E_{\mathcal{P}}|w_i(h) - E_{\mathcal{P}}[w_i(h)]|^{2+\eta} = o\left(\left[n^2\left(\frac{N-1}{N}\right)\Gamma_{\mathcal{P}}\right]^{(2+\eta)/2}\right). \tag{2.50}$$

Note that,

$$\begin{aligned}E_{\mathcal{P}}[w_i(h)] &= \frac{1}{N}\sum_{i\in U} w_i(h) \\ &= \frac{1}{N}\sum_{i\in U}[u_i(h) - v_i(h) + \frac{1}{N}\sum_{i\in U} v_j(h)] \\ &= \frac{1}{N}\sum_{i\in U} u_i(h) = \frac{1}{N}\sum_{i\in U} K_h(y-y_i) = f_U(y;h).\end{aligned}$$

Therefore,

$$\begin{aligned}
|w_i(h) - E_{\mathcal{P}}[w_i(h)]| &= \left| u_i(h) - v_i(h) + \frac{1}{N}\sum_{i\in U} v_j(h) - \frac{1}{N}\sum_{i\in U} u_i(h) \right| \\
&= \left| [u_i(h) - v_i(h)] - \frac{1}{N}\sum_{j\in U} [u_j(h) - v_j(h)] \right| \\
&\leq \left| \left[\frac{M}{h} - 0\right] - \frac{1}{N}\sum_{i\in U}\left[0 - \frac{M}{h}\right] \right| = \frac{2M}{h}.
\end{aligned}$$

$$\begin{aligned}
\sum_{i\in s} E_{\mathcal{P}} |w_i(h) - E_{\mathcal{P}}[w_i(h)]|^3 &\leq \sum_{i\in s} E_{\mathcal{P}} |w_i(h) - E_{\mathcal{P}}[w_i(h)]|^2 \max_{i\in s} |w_i(h) - E_{\mathcal{P}}[w_i(h)]| \\
&= \frac{2M}{h}\sum_{i\in s} E_{\mathcal{P}} [w_i(h) - E_{\mathcal{P}}\{w_i(h)\}]^2 \\
&= \frac{2M}{h}\sum_{i\in s} E_{\mathcal{P}} \left[\{u_i(h) - v_i(h)\} - \frac{1}{N}\sum_{j\in U}\{u_j(h) - v_j(h)\} \right]^2 \\
&= \frac{2M}{h}\frac{n}{N}\sum_{i\in U} \left[\{u_i(h) - v_i(h)\} - \frac{1}{N}\sum_{j\in U}\{u_j(h) - v_j(h)\} \right]^2 \\
&= \frac{2M}{h} n^2 \left(\frac{N-1}{N}\right)\left(1 - \frac{n}{N}\right)^{-1} \Gamma_{\mathcal{P}}.
\end{aligned}$$

Thus,

$$\left(1 - \frac{n}{N}\right)\sum_{i\in s} E_{\mathcal{P}} |w_i(h) - E_{\mathcal{P}}[w_i(h)]|^3 \leq \frac{2M}{h} n^2 \left(\frac{N-1}{N}\right)\Gamma_{\mathcal{P}}, \tag{2.51}$$

and

$$\begin{aligned}
&\frac{\left(1 - \frac{n}{N}\right)\sum_{i\in s} E_{\mathcal{P}} |w_i(h) - E_{\mathcal{P}}[w_i(h)]|^3}{\left[n^2\left(\frac{N-1}{N}\right)\Gamma_{\mathcal{P}}\right]^{3/2}} \\
&\leq \frac{2M}{h} n^2 \left(\frac{N-1}{N}\right)\Gamma_{\mathcal{P}} \left[n^2\left(\frac{N-1}{N}\right)\Gamma_{\mathcal{P}}\right]^{-3/2} \\
&= \frac{2M}{nh}\left(\frac{N-1}{N}\right)^{-1/2} \Gamma_{\mathcal{P}}^{-1/2} \\
&= \frac{2M}{\sqrt{n}h}\left(1 - \frac{n}{N}\right)^{-1/2} \left[\frac{\sum_{i\in U}(u_i(h) - v_i(h))^2 - N^{-1}\{\sum_{i\in U}(u_i(h) - v_i(h))\}^2}{N}\right]^{-1/2} \\
&\to 0, \qquad \text{as } \sqrt{n}h \to \infty.
\end{aligned} \tag{2.52}$$

Therefore, Lyapunov's condition (2.50) holds with $\eta = 1$ and the asymptotic result in (2.49) is proven. Since by Lemma 2.1, $\hat{f}_{dl}(y;h)$ and $\tilde{f}_{dl}(y;h)$ have the same limiting distribution in the design space, it follows from (2.49) that

$$\frac{\hat{f}_{dl}(y;h) - f_U(y;h)}{\Gamma_{\mathcal{P}}^{1/2}} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1). \tag{2.53}$$

To complete the proof of the lemma, it remains to show that $\hat{\Gamma}_{\mathcal{P}}$ is a design-consistent estimator for $\Gamma_{\mathcal{P}}$, or equivalently,

$$|\hat{\Gamma}_{\mathcal{P}} - \Gamma_{\mathcal{P}}| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.}$$

Consider using

$$\tilde{\sigma}_{\mathcal{P}}^2 = \frac{1}{n-1}\sum_{i\in s}\left[(u_i^*(h) - v_i^*(h)) - n^{-1}\{\sum_{i\in s}(u_i^*(h) - v_i^*(h))\}\right]^2 \tag{2.54}$$

as an estimator for

$$\sigma_{\mathcal{P}}^2 = \frac{1}{N}\sum_{i\in U}\left[(u_i^*(h) - v_i^*(h)) - N^{-1}\{\sum_{i\in U}(u_i^*(h) - v_i^*(h))\}\right]^2, \tag{2.55}$$

where $u_i^*(h) = hu_i(h) = K(\{y-y_i\}/h)$ and $v_i^*(h) = hv_i(h) = K(\{y-\beta_U x_i\}/h)$. Note that by the boundedness assumption on $K(\cdot)$,

$$0 < \sigma_{\mathcal{P}}^2 \leq \frac{1}{N}\sum_{i\in U}\left[(M-0) - N^{-1}\{\sum_{i\in U}(0-M)\}\right]^2 = 4M^2 < \infty. \tag{2.56}$$

Similarly,

$$\max_{i\in U}\left[(u_i^*(h) - v_i^*(h)) - N^{-1}\{\sum_{i\in U}(u_i^*(h) - v_i^*(h))\}\right]^2 / N\sigma_{\mathcal{P}}^2 \leq \frac{4M^2}{N\sigma_{\mathcal{P}}^2} \to 0 \text{ as } N \to \infty. \tag{2.57}$$

The bounds in (2.56) and (2.57) imply that conditions (2.10) and (2.12) on page 294 of Sen

(1988) are satisfied. Thus, using the result in (2.19) of Sen (1988), we have

$$|\tilde{\sigma}^2_{\mathcal{P}} - \sigma^2_{\mathcal{P}}| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.} \tag{2.58}$$

Consequently, taking $\tilde{\Gamma}_{\mathcal{P}} = (1-n/N)\tilde{\sigma}^2_{\mathcal{P}}/nh^2$, we have

$$|\tilde{\Gamma}_{\mathcal{P}} - \Gamma_{\mathcal{P}}| = \frac{1}{nh^2}\left(1-\frac{n}{N}\right)\left|\tilde{\sigma}^2_{\mathcal{P}} - \frac{N}{(N-1)}\sigma^2_{\mathcal{P}}\right| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.} \tag{2.59}$$

Define

$$\hat{\sigma}^2_{\mathcal{P}} = \frac{1}{n-1}\sum_{i\in s}\left[(u_i^*(h) - z_i^*(h)) - n^{-1}\{\sum_{i\in s}(u_i^*(h) - z_i^*(h))\}\right]^2,$$

where $z_i^*(h) = hz_i(h) = K(\{y - \hat{\beta}x_i\}/h)$, and observe that the only difference between $\hat{\sigma}^2_{\mathcal{P}}$ and $\tilde{\sigma}^2_{\mathcal{P}}$ is that $\hat{\sigma}^2_{\mathcal{P}}$ depends on $\hat{\beta}$ through $z_i^*(h)$ while $\tilde{\sigma}^2_{\mathcal{P}}$ depends on β_U through $v_i^*(h)$. Since, in the design space, $\hat{\beta}$ consistently estimates β_U, as we discussed right above Lemma 2.2, the results of Randles (1982) imply that $\hat{\sigma}^2_{\mathcal{P}}$ and $\tilde{\sigma}^2_{\mathcal{P}}$ share the same limiting properties. Additionally, it is readily seen that $\hat{\Gamma}_{\mathcal{P}} = (1-n/N)\hat{\sigma}^2_{\mathcal{P}}/nh^2$. Therefore, (2.59) implies that

$$|\hat{\Gamma}_{\mathcal{P}} - \Gamma_{\mathcal{P}}| = \frac{1}{nh^2}\left(1-\frac{n}{N}\right)\left|\hat{\sigma}^2_{\mathcal{P}} - \frac{N}{(N-1)}\sigma^2_{\mathcal{P}}\right| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.} \tag{2.60}$$

The proof of the lemma is complete upon using the results in (2.49) and (2.60) in conjunction with Slutsky's theorem. □

The following lemma is used to prove Theorem 2.3 which gives the asymptotic distribution of the estimator $\hat{f}_{dl}(y;h)$ under the combined mode of inference.

Lemma 2.4. *Suppose Assumptions A.1(i), A.2(i–ii) and A.3 hold. Then,*

$$\sqrt{Nh}[f_U(y;h) - E_\xi\{f_U(y;h)\}] \xrightarrow{\mathscr{L}_\xi} N(0, f(y)d_K), \tag{2.61}$$

where $\mathscr{L}_\xi$ means "convergence in law in the model space $(\Omega, \mathscr{F}, P_\xi)$".

Proof: The proof is identical to that of Parzen (1962, pg. 1069) due to the fact that $f_U(y;h)$ is the standard kernel density estimator from an IID sample where the sample is the entire finite population. □

Theorem 2.3. *Suppose all assumptions of Lemma 2.3 are satisfied. Then, under SRSWOR, we have*

$$\frac{\hat{f}_{dl}(y;h) - E_\xi\{f_U(y;h)\}}{\sqrt{V_C\{\hat{f}_{dl}(y;h)\}}} \xrightarrow{\mathscr{L}_C} N(0,1) \tag{2.62}$$

where $E_\xi\{f_U(y;h)\}$ *is given by* (2.20),

$$\begin{aligned} V_C\left[\hat{f}_{dl}(y;h)\right] &= \left(1-\frac{n}{N}\right)\frac{1}{nh^3}\left[\left(\frac{N-2}{N}\right)\frac{\mu_{XY}^2}{\mu_{X^2}^2}\int_{\mathbb{R}} x^2 t(x|y)dx - 2\left(\frac{N-1}{N}\right)\times \right. \\ &\qquad \left. \frac{\mu_{XY}}{\mu_{X^2}} y \int_{\mathbb{R}} xt(x|y)dx + y^2\right] f(y)d_{K'} + o\left(\frac{1}{Nh^3}\right) \\ &= \left(1-\frac{n}{N}\right)\frac{1}{nh^3}\varphi(x,y)f(y)d_{K'} + o\left(\frac{1}{Nh^3}\right) \end{aligned} \tag{2.63}$$

and $\mathscr{L}_C$ *means "convergence in law in the product space* $(\Omega \times S, \mathscr{F} \otimes \mathscr{S}, P_C)$*".*

Proof: First, note that

$$\begin{aligned} \sqrt{nh^3}[\hat{f}_{dl}(y;h) - E_\xi\{f_U(y;h)\}] &= \sqrt{nh^3}[\hat{f}_{dl}(y;h) - f_U(y;h)] \\ &\quad + \sqrt{nh^3}[f_U(y;h) - E_\xi\{f_U(y;h)\}]. \end{aligned} \tag{2.64}$$

Now, we have

$$\sqrt{nh^3}[f_U(y;h) - E_\xi\{f_U(y;h)\}] = h\sqrt{\frac{n}{N}}\sqrt{Nh}[f_U(y;h) - E_\xi\{f_U(y;h)\}] \xrightarrow{P_\xi} 0, \tag{2.65}$$

by using the assumptions that $h \to 0$ and $n/N \to \pi$ and Lemma 2.4 together in Slutsky's theorem. Moreover, from (2.53), $[\hat{f}_{dl}(y;h) - f_U(y;h)]\Gamma_{\mathcal{P}}^{-1/2}$ converges to the standard normal distribution in the design space where $\Gamma_{\mathcal{P}}$ is defined in (2.48). Using (2.48) and (2.55), we write $\Gamma_{\mathcal{P}} = (1-n/N)[N/(N-1)]\sigma_{\mathcal{P}}^2/nh^2$. Using results from the proof of Theorem 2.1, see

Eqs. (2.33) and (2.40), it can be shown that

$$E_\xi\left[\frac{1}{nh^2}\sigma^2_{\mathcal{P}}\right] = \frac{1}{nh^3}\varphi(x,y)f(y)d_{K'} + o\left(\frac{1}{Nh^3}\right).$$

Thus,

$$\begin{aligned}E_\xi(\Gamma_{\mathcal{P}}) &= \left(1-\frac{n}{N}\right)\frac{N}{(N-1)}\frac{1}{nh^3}\varphi(x,y)f(y)d_{K'} + o\left(\frac{1}{Nh^3}\right)\\ &= \frac{N}{(N-1)}V_C\left[\hat{f}_{dl}(y;\hat{\beta},h)\right].\end{aligned}$$

Applying Theorem 5.1 of Bleuer and Kratina (2005) and Slutsky's theorem completes the proof. □

2.5. Bandwidth Selection

The fact that the bandwidth has a crucial effect on the performance of kernel density estimators in general is present for the proposed estimator $\hat{f}_{dl}(y;h)$ as is seen from the MISE formula given in Theorem 2.2. From the MISE formula in (2.17), we see that small values for the bandwidth h, make the estimator very variable (under-smoothed) and less biased while large values for h result in an over-smoothed highly biased estimator which fails to reveal structural features such as multi-modality. Thus, methods for selecting the optimal bandwidth, in the sense that it makes a reasonable balance between the bias and the variance of the proposed estimator, are required. There exist several bandwidth selection methods in the literature on kernel density estimation. Some of these methods have been reviewed in Section 1.2.3. In this section, the direct plug-in method is used to select the amount of smoothing for the proposed model-assisted estimator $\hat{f}_{dl}(y;h)$.

From Theorem 2.2, the AMISE of the estimator $\hat{f}_{dl}(y;h)$ is given by

$$AMISE\left[\hat{f}_{dl}(\cdot;h)\right] = \frac{1}{nh^3}\delta\mu_{Y^2}\left(1-\frac{\mu^2_{XY}}{\mu_{X^2}\mu_{Y^2}}\right)d_{K'} + \frac{1}{4}h^4c^2_K d_{f''}. \tag{2.66}$$

Using this formula for the AMISE, we derive the asymptotically optimal global bandwidth $h_{opt,dl}$ as follows. First, note that

$$\frac{\partial}{\partial h} AMISE\left[\hat{f}_{dl}(\cdot;h)\right] = -3\delta \frac{1}{nh^4}\mu_{_{Y2}}\left(1-\frac{\mu_{_{XY}}^2}{\mu_{_{X2}}\mu_{_{Y2}}}\right)d_{K'} + h^3 c_K^2 d_{f''}.$$

Setting the above partial equal to zero and solving for h, we get

$$h_{opt,dl} = \left[3\delta\mu_{_{Y2}}\left(1-\frac{\mu_{_{XY}}^2}{\mu_{_{X2}}\mu_{_{Y2}}}\right)d_{K'}\right]^{1/7}\left(c_K^2 d_{f''}\right)^{-1/7} n^{-1/7}. \tag{2.67}$$

Note that the optimal bandwidth $h_{opt,dl}$ is $O(n^{-1/7})$ which is larger than $O(n^{-1/5})$, the order of the bandwidth for standard kernel density estimators with no auxiliary data. Again, this is due to the bivariate setting imposed by the auxiliary data.

Clearly, the optimal bandwidth in (2.67) is not obtainable as it depends on the unknown quantities; $\mu_{_{X2}}$, $\mu_{_{Y2}}$, $\mu_{_{XY}}^2$ and $d_{f''} \equiv \int\{f''(y)\}^2 dy$. The idea of direct plug-in methods, described briefly in Section 1.2.3, is based on plugging in estimates for these unknown quantities. Noting that the first three quantities are moments of the densities $g(x)$, $f(y)$ and $t(x,y)$, respectively, we can use the following corresponding sample moments to estimate these quantities:

$$\hat{\mu}_{_{X2}} = \frac{1}{n}\sum_{i\in s} x_i^2, \quad \hat{\mu}_{_{Y2}} = \frac{1}{n}\sum_{i\in s} y_i^2, \quad \text{and} \quad \hat{\mu}_{_{XY}} = \frac{1}{n}\sum_{i\in s} x_i y_i. \tag{2.68}$$

It remains to estimate the functional $d_{f''}$. There exists huge literature on estimation of density functionals (e.g., Wand and Jones (1995, pg. 67)). Using integration by parts, it can be shown that for an s-order functional,

$$d_{f^{(s)}} = (-1)^s \int f^{(2s)}(y) f(y) dy. \tag{2.69}$$

Therefore, to estimate $d_{f^{(s)}}$, it suffices to estimate

$$\Psi_r = \int f^{(r)}(y) f(y) dy,$$

for r even. Noting that $\Psi_r = E_\xi[f^{(r)}(Y)]$, Hall and Marron (1987) suggested estimating Ψ_r by the estimator

$$\hat{\Psi}_r(b_1) = \frac{1}{n} \sum_{i \in s} \hat{f}^{(r)}(y_i; b_1), \tag{2.70}$$

where $\hat{f}^{(r)}(y_i; b_1) = n^{-1} \sum_{j \in s} W_{b_1}^{(r)}(y_i - y_j)$ with W and b_1 being a kernel function and a bandwidth, respectively. Since, from (2.69), $d_{f''} = \Psi_4$, we can estimate $d_{f''}$ using $\hat{\Psi}_4(b_1)$. However, the estimate $\hat{\Psi}_4(b_1)$ depends on the choice of the pilot bandwidth b_1. The AMISE-optimal b_1 for the estimate $\hat{\Psi}_4(b_1)$ takes the following form (e.g., Wand and Jones (1995, pg. 71)):

$$b_{1,opt} = \left[-2\chi^{(4)}(0)/\{c_W \Psi_6 n\}\right]^{1/7}.$$

Note that $b_{1,opt}$ depends on the unknown functional Ψ_6 which can be estimated by $\hat{\Psi}_6(b_2)$ which requires the selection of a new bandwidth b_2. It is clear that this process can continue for ever. This problem can be tackled by choosing the number of stages of functional estimation to be performed and then using a quick and simple method to select the bandwidth needed for the last stage, say in $\hat{\Psi}_l(w)$. Sheather and Jones (1991) recommended using two-stage plug-in bandwidth selectors. Taking $W = K$, the same second-order kernel, two-stage plug-in bandwidth selectors involve four steps:

Step 1. Estimate Ψ_8 using $\hat{\Psi}_8(b_3)$ which is defined in (2.70) where $b_3 = 105/(32\hat{\sigma}^9\sqrt{\pi})$ is the normal rule-of-thumb bandwidth with $\hat{\sigma}$ being any scale estimate.

Step 2. Estimate Ψ_6 using $\hat{\Psi}_6(b_2)$ where $b_2 = [-2K^{(6)}(0)/\{c_K \hat{\Psi}_8(b_3) n\}]^{1/9}$.

Step 3. Estimate Ψ_4 using $\hat{\Psi}_4(b_1)$ where $b_1 = [-2K^{(4)}(0)/\{c_K \hat{\Psi}_6(b_2) n\}]^{1/7}$.

Step 4. A two-stage direct plug-in estimator of h is

$$\hat{h}_{DPI,dl} = \left[3\delta\hat{\mu}_{Y^2}\left(1 - \frac{\hat{\mu}_{XY}^2}{\hat{\mu}_{X^2}\hat{\mu}_{Y^2}}\right) d_{K'}\right]^{1/7} \left[c_K^2 \hat{\Psi}_4(b_1) n\right]^{-1/7}. \tag{2.71}$$

This bandwidth estimator is used to compute the estimator $\hat{f}_{dl}(y;h)$ in the Mote Carlo study of Chapter 5.

CHAPTER 3

KDE Using Auxiliary Information Via Nonparametric Regression Models of the Study Variable

3.1. Introduction

In Chapter 2, auxiliary data were utilized in estimating the density function of the study variable Y via modeling the relationship between X and Y using parametric linear regression models. It will be demonstrated from the simulation study in Chapter 5 that the model-assisted kernel density estimator we proposed in Chapter 2 performs well when the linear relationship exists. However, the assumption of having a linear (or any other parametric) relationship between X and Y may be two rigid in some situations. In this chapter, we relax this assumption to account for the case where the relationship between the two variables is unknown and we cannot assume a certain form of relationship to hold. In the later case, the relationship between the two variables can be best described using nonparametric regression models. These models are described in Section 3.2. In the same section, we propose another model-assisted kernel density estimator for $f_Y(\cdot)$. The asymptotic properties of the new estimator are studied in Sections 3.3 and 3.4. Section 3.5 deals with the bandwidth selection problem for the proposed estimator.

3.2. Model-Assisted KDE Using Nonparametric Regression Models for Y on X

Suppose the relationship between X and Y can be written in the following form:

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{3.1}$$

where $m(x) = E_\xi(Y|X = x)$ and $\sigma^2(x) = V_\xi(Y|X = x)$ are unspecified smooth functions and the variables ε_i are IID with zero mean and unit variance. The model in (3.1) is called the random design regression model because (X, Y) are assumed to be observed as a bivariate sample. Several techniques can be used for fitting this regression model. These techniques include kernel regression, splines and wavelets. Among these techniques, kernel regression is most popular due to being mathematically and intuitively simple. In this dissertation, we adopt the kernel regression technique, more specifically local polynomial kernel estimators, to fit the regression model in (3.1). The resulting fits are then used to define the second model-assisted kernel density estimator for the density of Y.

Local polynomial kernel estimators estimate the regression function $m(\cdot)$ at a certain point by locally fitting a p-th degree polynomial to the data using a kernel-weighted least squares method. Wand and Jones (1995, Ch. 5) give a good introduction to these estimators. If values of both X and Y are known for the entire finite population, the p-th degree local polynomial estimator of the regression function $m(\cdot)$ at the point x_i is given by

$$m_U(x_i; p, a) = \mathbf{e}_1^T \left[\mathbf{X}_{iU}^T \mathbf{W}_{iU} \mathbf{X}_{iU}\right]^{-1} \mathbf{X}_{iU}^T \mathbf{W}_{iU} \mathbf{y}_U, \tag{3.2}$$

where $\mathbf{X}_{iU}$ is an $N \times (p+1)$ design matrix having the following form

$$\mathbf{X}_{iU} = \begin{bmatrix} 1 & (X_1 - x_i) & \ldots & (X_1 - x_i)^p \\ 1 & (X_2 - x_i) & \ldots & (X_2 - x_i)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_N - x_i) & \ldots & (X_N - x_i)^p \end{bmatrix} = \left[1 \ (X_j - x_i) \ \cdots \ (X_j - x_i)^p\right]_{j \in U},$$

$\mathbf{W}_{iU}$ is an $N \times N$ weighting matrix of the form

$$\mathbf{W}_{iU} = diag\left\{\psi_a(X_j - x_i)\right\}_{j\in U},$$

where ψ is a kernel function and a is a bandwidth sequence, $\mathbf{y}_U = (Y_1\ Y_2\ \ldots\ Y_N)^T$ and $\mathbf{e}_r$ is a $(p+1)\times 1$ vector with 1 in the r-th entry and zero elsewhere. Such one non-null entry vectors will be used extensively in this dissertation and the dimension will be clear from the context. Note that the estimator $m_U(\cdot; p, a)$ is not obtainable as it requires observing the study variable Y for the entire finite population while it is only observable for the sample. The design-based local polynomial estimator of $m(\cdot)$ at the point x_i has the form

$$\hat{m}(x_i; p, a) = \mathbf{e}_1^T \left[\mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{X}_{is}\right]^{-1} \mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{y}_s, \tag{3.3}$$

where $\mathbf{X}_{is}$ is a $n\times(p+1)$ sample design matrix having the form

$$\mathbf{X}_{is} = \left[1\ (x_j - x_i)\ \cdots\ (x_j - x_i)^p\right]_{j\in s},$$

$\mathbf{W}_{is} = diag\left\{d_j\psi_a(x_j - x_i)\right\}_{j\in s}$ is an $n\times n$ sample weighting matrix with d_j being the sampling weights and $\mathbf{y}_s$ is the sample analog of $\mathbf{y}_U$. Two special cases of $m_U(\cdot; p, a)$ and $\hat{m}(\cdot; p, a)$ are obtained when we take $p = 0, 1$. When $p = 0$, the finite population fit $m_U(\cdot; 0, a)$ reduces to the well-known Nadaraya-Watson regression estimator (see, Nadaraya (1964) and Watson (1964));

$$m_U(x_o; 0, a) = \frac{\sum_{j\in U}\psi_a(X_j - x_o)Y_j}{\sum_{j\in U}\psi_a(X_j - x_o)}$$

and the sample-based fit $\hat{m}(\cdot; 0, a)$ reduces to

$$\hat{m}(x_o; 0, a) = \frac{\sum_{j\in s} d_j\psi_a(x_j - x_o)y_j}{\sum_{j\in s} d_j\psi_a(x_j - x_o)}. \tag{3.4}$$

For $p = 1$, we get the following local linear estimators:

$$m_U(x_o; 1, a) = \sum_{j \in U} \frac{\{t_2(x_o; a) - (X_j - x_o)t_1(x_o; a)\}\psi_a(X_j - x_o)Y_j}{t_2(x_o; a)t_0(x_o; a) - t_1^2(x_o; a)},$$

and

$$\hat{m}(x_o; 1, a) = \sum_{j \in s} \frac{d_j\{\hat{t}_2(x_o; a) - (x_j - x_o)\hat{t}_1(x_o; a)\}\psi_a(x_j - x_o)y_j}{\hat{t}_2(x_o; a)\hat{t}_0(x_o; a) - \hat{t}_1^2(x_o; a)},$$

where $t_r(x_o; a) = \sum_{k \in U}(X_k - x_o)^r \psi_a(X_k - x_o)$ and $\hat{t}_r(x_o; a) = \sum_{k \in s} d_k(x_k - x_o)^r \psi_a(x_k - x_o)$ for $r = 0, 1, 2$.

An important issue that arise when using the local polynomial regression technique is the choice of the polynomial degree, p. Wand and Jones (1995, pg. 126) indicate that, although choosing higher values for p improves the asymptotic performance of $\hat{m}(x_o; p, a)$, the practical gains are not as clear-cut. They recommend using odd degree fits, specially $p = 1$ or 3 because of their attractive bias formulae and boundary properties. For the same reason and since the smoothing parameter, a, controls the modeling complexity, Fan and Gijbels (1996, pg. 59) recommend the use of the lowest odd order, i.e., $p = 1$, or occasionally $p = 3$. Below, we define a new model-assisted kernel density estimator for the density of the study variable, $f(\cdot)$, using the general p-th degree local polynomial regression estimate $\hat{m}(\cdot; p, a)$ which is given in Eq. (3.3). In the rest of this chapter, we study the properties of the proposed density estimator in its general form. Only in Theorems 3.2 and 3.3, we restrict our analysis to the case where $\hat{m}(\cdot; 0, a)$, the Nadaraya-Watson regression estimator, is used to estimate the regression function $m(\cdot)$. Using this specific regression estimator keeps the derivation of the asymptotic expression of the MSE of the density estimator, which we propose in the next section, tractable. If the local constant estimate $\hat{m}(\cdot; 0, a)$ is replaced by the local linear estimate $\hat{m}(\cdot; 1, a)$ in Theorem 3.2, a more complicated argument, than the one we use in the proof of the theorem, would be required to obtain the asymptotic expression of the MSE under the combined mode of inference.

3.2.1. Proposed Estimator

Using the available auxiliary data, we can obtain the fitted values $\hat{Y}_i = \hat{m}(X_i)$ for $i \in U$, which, in turn, can be used to define the following model-assisted kernel density estimator for the density $f(y)$:

$$\hat{f}_{dn}(y;h) = \frac{1}{N}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i) - K_h(y-\hat{m}(x_i))\right\} + \sum_{i\in U} K_h(y-\hat{m}(X_i))\right], \tag{3.5}$$

where $\hat{m}(x_i) = \hat{m}(x_i;p,a)$ is as in (3.3). For ease of notation, sometimes, we will use $\hat{m}_i$ and m_{Ui} to refer to $\hat{m}(X_i;p,a)$ and $m_U(X_i;p,a)$, respectively. The estimator $\hat{f}_{dn}(\cdot;h)$ enjoys two nice properties; (1) it reveals the restriction of parametric models usually used to define model-assisted estimators; (2) it incorporates the design information through using the design weights which causes the estimator to be both asymptotically unbiased and consistent with respect to the randomization distribution.

3.2.2. Main Assumptions

In addition to Assumptions A.1–A.4 in Section 2.2.2, we need the following conditions to study the statistical properties of the estimator in (3.5).

B.1 *(The kernel function ψ)*:

(i) The function $\psi(\cdot)$ is a bounded density function that is symmetric about the origin.

(ii) $\int z\psi^2(z)dz = 0$, $\int z^2\psi(z)dz = c_\psi < \infty$ and $\int \psi^2(z)dz = d_\psi < \infty$.

B.2 *(The bandwidth a)*: $a_\tau(n_\tau, N_\tau) = a_\tau = a$ is such that $a_\tau \to 0$, $n_\tau a_\tau \to \infty$ and $h_\tau = O(a_\tau)$ as $\tau \to \infty$.

B.3 *(Regression and variance functions)*:

(i) The regression function $m(x) = E_\xi(Y|X=x)$ has a bounded continuous second derivative.

(ii) The variance function $\sigma^2(x) = V_\xi(Y|X=x)$ is bounded and continuous.

3.3. Properties of $\hat{f}_{dn}(y)$

The following are three basic properties of the estimator $\hat{f}_{dn}(y)$:

Property 1. Under any sampling design, the estimator $\hat{f}_{dn}(y)$, defined in (3.5), is a true probability density function with probability approaching 1.

Proof: First,

$$\int_{\mathbb{R}} \hat{f}_{dn}(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i \left\{ \int_{\mathbb{R}} K_h(y-y_i)dy - \int_{\mathbb{R}} K_h(y-\hat{m}_i)dy \right\} + \sum_{i\in U} \int_{\mathbb{R}} K_h(y-\hat{m}_i)dy \right].$$

Using the change of variables $z = (y-u)/h$, we have

$$\int_{\mathbb{R}} \hat{f}_{dn}(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i \left\{ \int_{\mathbb{R}} K(z)dz - \int_{\mathbb{R}} K(z)dz \right\} + \sum_{i\in U} \int_{\mathbb{R}} K(z)dz \right] = 1.$$

Second, $\hat{f}_{dn}(y) \geq 0$ for all $y \in \mathbb{R}$ with probability approaching 1 since $\hat{f}_{dn}(y)$ is a consistent estimator, in the MSE sense, for $f(y)$ (see the MISE expression in Theorem 3.2). □

Property 2. The mean of the estimator $\hat{f}_{dn}(y)$ is the local polynomial model-assisted estimator of the finite population mean $\bar{Y}_U$ (e.g., Breidt and Opsomer (2000)):

$$\hat{\bar{Y}}_{U,LP} = \int_{-\infty}^{\infty} y\hat{f}_{dn}(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i(y_i-\hat{m}_i) + \sum_{i\in U} \hat{m}_i \right]. \tag{3.6}$$

If, instead of the local polynomial regression technique, we use the technique of penalized splines to fit the model in (3.1), the mean of the new density estimator will resemble the penalized spline model-assisted estimator of $\bar{Y}_U$ due to Breidt et al. (2005).

Proof:

$$\begin{aligned}
&\int_{\mathbb{R}} y\hat{f}_{dn}(y)dy \\
&= \frac{1}{N}\left[\sum_{i\in s} d_i \int_{\mathbb{R}} yK_h(y-y_i)dy - \sum_{i\in s} d_i \int_{\mathbb{R}} yK_h(y-\hat{m}_i)dy + \sum_{i\in U} \int_{\mathbb{R}} yK_h(y-\hat{m}_i)dy \right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\left[\sum_{i\in s} d_i \int_{\mathbb{R}} (y_i + hz)K(z)dz - \sum_{i\in s} d_i \int_{\mathbb{R}} (\hat{m}_i + hz)K(z)dz + \sum_{i\in U} \int_{\mathbb{R}} (\hat{m}_i + hz)K(z)dz\right] \\
&= \frac{1}{N}\left[\sum_{i\in s} d_i y_i - \sum_{i\in s} d_i \hat{m}_i + \sum_{i\in U} \hat{m}_i\right] \\
&= \frac{1}{N}\sum_{i\in s} d_i (y_i - \hat{m}_i) + \frac{1}{N}\sum_{i\in U} \hat{m}_i,
\end{aligned}
$$

where the second and third qualities follow from using the change of variables $z = (y-u)/h$ and from Assumption A.2(ii), respectively. □

Property 3. Integrating the estimator $\hat{f}_{dn}(y)$ over the interval $(-\infty, t]$ gives a smooth model-assisted estimator of the distribution function $F(t)$. This estimator can be viewed as a smooth version of Dorfman and Hall (1993) model-assisted estimator of the finite population CDF, $F_U(t)$ (see Eq. (1.27) in Section 1.5):

$$
\begin{aligned}
\hat{F}_{GDNS}(t) &= \int_{-\infty}^{t} \hat{f}_{dn}(y)dy \\
&= \frac{1}{N}\left\{\sum_{i\in U} \mathcal{K}\left(\frac{t-\hat{m}(X_i)}{h}\right) + \sum_{i\in s} d_i \mathcal{K}\left(\frac{t-y_i}{h}\right) - \sum_{i\in s} d_i \mathcal{K}\left(\frac{t-\hat{m}(x_i)}{h}\right)\right\},
\end{aligned}
$$

where $\mathcal{K}(y) = \int_{-\infty}^{y} K(u)du$.

Proof: The proof is similar to that of Property 3 in Section 2.3. □

Note that the density estimator $\hat{f}_{dn}(\cdot;h)$ in (3.5) involves the sample-based local polynomial regression estimate $\hat{m}(\cdot)$, defined in (3.3), which complicates the derivation of many of the properties of the density estimator itself. To overcome this issue, we use results from Randles (1982) to show that the limiting design-based distribution of the estimator $\hat{f}_{dn}(\cdot;h)$ remains the same if we replace $\hat{m}(\cdot)$ with the finite population fit $m_U(\cdot)$, defined in (3.2). That is, the limiting design-based distributions of the following pseudo density estimator

$$
\tilde{f}_{dn}(y;h) = \frac{1}{N}\left[\sum_{i\in s} d_i \left\{K_h(y-y_i) - K_h(y-m_U(x_i))\right\} + \sum_{i\in U} K_h(y - m_U(X_i))\right], \qquad (3.7)
$$

where $m_U(x_i) = m_U(x_i; p, a)$ is as in (3.2), and the proposed estimator $\hat{f}_{dn}(\cdot;h)$ are identical. This observation is formulated in the following lemma.

Lemma 3.1. *Under the smoothness assumptions of Section 2.2.2 on the kernel $K(\cdot)$, the two estimators $\hat{f}_{dn}(y;h)$ and $\tilde{f}_{dn}(y;h)$ have the same limiting design-based distribution.*

Proof: Follows the same lines of the proof of Lemma 2.2. □

The design-based properties of the estimator $\hat{f}_{dn}(y;h)$ are summarized in the following theorem.

Theorem 3.1. *Suppose Assumptions A.3 and A.4 in Section 2.2.2 hold and $K(x) \leq M$ for all x. Then, the estimator $\hat{f}_{dn}(y;h)$ is asymptotically design-unbiased and design-consistent (in the MSE sense) for $f_U(y)$.*

Proof: Follows the same lines of the proof of Theorem 2.1. □

The following theorem gives the bias and MISE formulae of the estimator $\hat{f}_{dl}(y;h)$ under the combined mode of inference.

Theorem 3.2. *Suppose Assumptions A.1–A.4 and B.1–B.3 hold. Then, under the combined mode of inference, the bias and the MISE of the estimator $\hat{f}_{dn}(y;h)$, when $\hat{m}(\cdot)$ in (3.5) is taken to be the Nadaraya-Watson estimator $\hat{m}(\cdot;0,a)$, are given by:*

$$Bias_C\left[\hat{f}_{dn}(y;h)\right] = \frac{1}{2}h^2 c_K f''(y) + o(h^2), \tag{3.8}$$

and

$$\begin{aligned} MISE_C\left[\hat{f}_{dn}(\cdot;h)\right] = & \frac{1}{nh^3}\delta\left[\frac{E_\xi\{m^2(X)g^2(X)\}}{S_g} - 2\frac{E_\xi\{m^2(X)g(X)\}}{d_g} + \mu_{Y^2}\right] d_{K'} \\ & + \frac{1}{4}h^4 c_K^2 d_{f''} + o\left(h^4 + \frac{1}{Nh^3}\right), \end{aligned} \tag{3.9}$$

where $\delta = nN^{-2}\sum_{i\in U}\Delta_i$, $d_g = \int g^2(u)du$ and $S_g = \int g^3(u)du$.

Remarks: The following are some noteworthy points to be taken from Theorem 3.2.

(i) Like the estimator $\hat{f}_{dl}(y)$ in Chapter 2, the leading term in the bias of the estimator $\hat{f}_{dn}(y)$ is identical to the leading term of the bias of the standard kernel density estimator that does not use any auxiliary information (see Section 1.2). This implies that the auxiliary information and the sampling scheme do not affect the leading term in the bias of kernel density estimators.

(ii) The effects of auxiliary data and design weights are reflected in the variance of $\hat{f}_{dn}(y)$ as can be seen from the first term in the MISE formula in (3.9). In this term, the quantity δ is an adjusting factor that results from using the design weights in the estimator $\hat{f}_{dn}(y)$. The quantities between the square brackets in the first term of (3.9) implicitly reflect the dependence of the asymptotic behavior of $\hat{f}_{dn}(y)$ on the correlation between X and Y. Simple algebraic manipulations show that this term is always nonnegative and is indeed bounded below by $E\{\sigma^2(X)\}$ where $\sigma^2(X)$ is the conditional variance function in model (3.1).

(iii) Unlike in standard kernel density estimators, the bandwidth h has a third power in the denominator of the asymptotic variance of $\hat{f}_{dn}(y;h)$. This is due to the bivariate setting created by incorporating the auxiliary information in the structure of $\hat{f}_{dn}(y;h)$.

Proof: Starting with the bias statement and using rules of conditional expectation, we have

$$E_C\left[\hat{f}_{dn}(y;h)\right] = E_\xi\left\{E_\mathcal{P}\left[\hat{f}_{dn}(y;h)|\mathbf{X}_U,\mathbf{Y}_U\right]\right\}.$$

Using Lemma 3.1, we have

$$\begin{aligned}
&E_\mathcal{P}\left[\hat{f}_{dn}(y;h)\right]\\
&\approx E_\mathcal{P}\left[\tilde{f}_{dn}(y;h)\right]\\
&= \frac{1}{N}E_\mathcal{P}\left[\sum_{i\in s} d_i\left\{K_h(y-y_i)-K_h(y-m_U(x_i))\right\}+\sum_{i\in U}K_h(y-m_U(X_i))\right]\\
&= \frac{1}{N}\left[\sum_{i\in U} E_\mathcal{P}(I_i)d_i\left\{K_h(y-Y_i)-K_h(y-m_U(X_i))\right\}+\sum_{i\in U}K_h(y-m_U(X_i))\right]\\
&= \frac{1}{N}\left[\sum_{i\in U}\left\{K_h(y-Y_i)-K_h(y-m_U(X_i))\right\}+\sum_{i\in U}K_h(y-m_U(X_i))\right]\\
&= \frac{1}{N}\sum_{i\in U}K_h(y-Y_i) = f_U(y;h).
\end{aligned} \tag{3.10}$$

Consequently,

$$E_C\left[\hat{f}_{dn}(y;h)\right]=\frac{1}{N}\sum_{i\in U}E_\xi\left[K_h(y-Y_i)\right]=f(y)+\frac{1}{2}h^2c_Kf''(y)+o(h^2), \tag{3.11}$$

where the second equality in (3.11) is a standard result in kernel density estimation (e.g., Wand and Jones (1995, pg. 20)). Subtracting $f(y)$ from the right hand side of 3.11 gives the bias of $\hat{f}_{nd}(y;h)$ as in (3.8).

Next, we work on the variance of the estimator $\hat{f}_{dn}(y;h)$. Again, using rules of conditional expectations, we can write

$$\begin{aligned}V_C\left[\hat{f}_{dn}(y;h)\right] &= E_\xi\left\{V_{\mathcal{P}}\left[\hat{f}_{dn}(y;h)|\mathbf{X}_U,\mathbf{Y}_U\right]\right\}+V_\xi\left\{E_{\mathcal{P}}\left[\hat{f}_{dn}(y;h)|\mathbf{X}_U,\mathbf{Y}_U\right]\right\}\\ &:= L_1+L_2. \end{aligned}\tag{3.12}$$

From (3.10), we have

$$\begin{aligned}L_2 &\approx V_\xi\left[f_U(y;h)\right]=V_\xi\left[\frac{1}{N}\sum_{i\in U}K_h(y-Y_i)\right]\\ &\overset{iid}{=}\frac{1}{N}V_\xi\left[K_h(y-Y_1)\right]=(Nh)^{-1}d_Kf(y)+o\{(Nh)^{-1}\},\end{aligned}\tag{3.13}$$

where the last equality is a standard result (e.g., Wand and Jones (1995, pg. 21)).

On the other hand, to evaluate the first term in (3.12), we start by finding the approximate design variance of $\hat{f}_{dn}(\cdot;h)$ as follows. From Lemma 2.3, we can write

$$\begin{aligned}&V_{\mathcal{P}}\left[\hat{f}_{dn}(y;h)\right]\\ &\quad\approx V_{\mathcal{P}}\left[\tilde{f}_{dl}(y;h)\right]\\ &\quad=\frac{1}{N^2}V_{\mathcal{P}}\left[\sum_{i\in s}d_i\left\{K_h(y-y_i)-K_h(y-m_{Ui})\right\}+\sum_{i\in U}K_h(y-m_{Ui})\right]\\ &\quad=\frac{1}{N^2}V_{\mathcal{P}}\left[\sum_{i\in s}d_i\left\{K_h(y-y_i)-K_h(y-m_{Ui})\right\}\right]\\ &\quad=\frac{1}{N^2}\sum\sum_{i,j\in U}\Delta_{ij}\left\{K_h(y-y_i)-K_h(y-m_{Ui})\right\}\left\{K_h(y-y_j)-K_h(y-m_{Uj})\right\}.\end{aligned}\tag{3.14}$$

To reach (3.14), we used the fact that the second summation in the first equality is a finite population quantity and hence considered fixed with respect to the sampling design so its variance is zero. The last equality follows because $\sum_{i\in s} d_i \{K_h(y-y_i)-K_h(y-m_{Ui})\}$ is the Horvitz-Thompson estimator (see Eq. (1.15) in Section 1.5) of the finite population total of the variable $\{K_h(y-y_i)-K_h(y-m_{Ui})\}$. Taking $\hat{Y}_{Ui}=m_U(X_i)=m_{Ui}$, (3.14) can be rewritten as follows;

$$
\begin{aligned}
&V_{\mathcal{P}}\left[\hat{f}_{dn}(y;h)\right] \\
&\quad = \frac{1}{N^2}\sum_{i\in U}\Delta_i\left\{K_h(y-Y_i)-K_h(y-\hat{Y}_{Ui})\right\}^2 \\
&\quad\quad +\frac{1}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij}\left\{K_h(y-Y_i)-K_h(y-\hat{Y}_{Ui})\right\}\left\{K_h(y-Y_j)-K_h(y-\hat{Y}_{Uj})\right\} \\
&\quad = \frac{1}{N^2}\sum_{i\in U}\Delta_i W_i^2+\frac{1}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij}W_iW_j. \qquad (3.15)
\end{aligned}
$$

Then, from (3.12) and (3.15) together, we can write

$$
L_1 \approx \frac{1}{N^2}\sum_{i\in U}\Delta_i E_\xi(W_i^2)+\frac{1}{N^2}\sum\sum_{i,j\in U, i\neq j}\Delta_{ij}E_\xi(W_iW_j). \qquad (3.16)
$$

We work each term in (3.16) separately. First, notice that by a Taylor expansion, we have

$$
K_h(\hat{Y}_{Ui}-y)\approx\frac{1}{h}K\left(\frac{Y_i-y}{h}\right)+\frac{(\hat{Y}_{Ui}-Y_i)}{h^2}K'\left(\frac{Y_i-y}{h}\right),
$$

and

$$
\begin{aligned}
K_h^2(\hat{Y}_{Ui}-y)\approx\ &\frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right)+2\frac{(\hat{Y}_{Ui}-Y_i)}{h^3}K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_i-y}{h}\right) \\
&+\frac{(\hat{Y}_{Ui}-Y_i)^2}{h^4}\left[K'\left(\frac{Y_i-y}{h}\right)\right]^2.
\end{aligned}
$$

Using these expansions, we can write

$$
W_i^2 \;=\; K_h^2(Y_i-y)-2K_h(Y_i-y)K_h(\hat{Y}_{Ui}-y)+K_h^2(\hat{Y}_{Ui}-y)
$$

$$
\begin{aligned}
&\approx \frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right)-2\left[\frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right)+\frac{(\hat{Y}_{Ui}-Y_i)}{h^3}K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_i-y}{h}\right)\right]\\
&\quad+\frac{1}{h^2}K^2\left(\frac{Y_i-y}{h}\right)+2\frac{(\hat{Y}_{Ui}-Y_i)}{h^3}K\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_i-y}{h}\right)\\
&\quad+\frac{(\hat{Y}_{Ui}-Y_i)^2}{h^4}K'^2\left(\frac{Y_i-y}{h}\right)\\
&= \frac{1}{h^4}\hat{Y}_{Ui}^2K'^2\left(\frac{Y_i-y}{h}\right)-\frac{2}{h^4}\hat{Y}_{Ui}Y_iK'^2\left(\frac{Y_i-y}{h}\right)+\frac{1}{h^4}Y_i^2K'^2\left(\frac{Y_i-y}{h}\right)\\
&:= M_1-2M_2+M_3. \qquad (3.17)
\end{aligned}
$$

We now evaluate the expectation of each term on the right-hand side of (3.17), in reverse.

$$
\begin{aligned}
\frac{1}{N}E_\xi(M_3) &= \frac{1}{Nh^4}\int_{\mathbb{R}} y_i^2K'^2\left(\frac{y_i-y}{h}\right)f(y_i)dy_i\\
&= \frac{1}{Nh^3}\int_{\mathbb{R}}(y+hz)^2K'^2(z)(y+hz)dz\\
&= \frac{1}{Nh^3}\int_{\mathbb{R}}K'^2(z)(y^2+2yhz+h^2z^2)\left[f(y)+hzf'(y)+\dots\right]dz\\
&= \frac{1}{Nh^3}y^2f(y)\int_{\mathbb{R}}K'^2(z)dz+\frac{1}{Nh^2}\left[y^2f'(y)+2yf(y)\right]\int_{\mathbb{R}}zK'^2(z)dz\\
&\quad+\frac{1}{Nh}\left[f(y)+2yf'(y)\right]\int_{\mathbb{R}}z^2K'^2(z)dz+o\left(\frac{1}{Nh}\right)\\
&= \frac{1}{Nh^3}y^2f(y)d_{K'}+\frac{1}{Nh}\left[f(y)+2yf'(y)\right]c^*_{K'}+o\left(\frac{1}{Nh}\right). \qquad (3.18)
\end{aligned}
$$

where (3.18) follows from the fact that $\int_{\mathbb{R}} zK'^2(z)dz = 0$ by assumptions on K. For M_2, observe that

$$
\begin{aligned}
\frac{1}{N}E_\xi(M_2) &= \frac{1}{Nh^4}E_\xi\left\{m_U(X_i)Y_iK'^2\left(\frac{Y_i-y}{h}\right)\right\}\\
&= \frac{1}{Nh^4}E_\xi\left\{\left(\frac{\sum_{j\in U}\psi_a(X_j-X_i)Y_j}{\sum_{j\in U}\psi_a(X_j-X_i)}\right)Y_iK'^2\left(\frac{Y_i-y}{h}\right)\right\}\\
&= \frac{1}{Nh^4}E_\xi\left\{\frac{\left(\sum_{j\in U}\psi_a(X_j-X_i)Y_j\right)Y_iK'^2\left(\frac{Y_i-y}{h}\right)}{\sum_{j\in U}\psi_a(X_j-X_i)}\right\}\\
&:= \frac{1}{Nh^4}E_\xi\left\{\frac{Q_1}{Q_2}\right\}\approx\frac{1}{Nh^4}\frac{E_\xi(Q_1)}{E_\xi(Q_2)}. \qquad (3.19)
\end{aligned}
$$

Clearly,

$$
\begin{aligned}
Q_1 \; &= \; \frac{1}{a}\psi(0)Y_i^2 K'^2\left(\frac{Y_i-y}{h}\right)+\frac{1}{a}\left(\sum_{j\in U, j\neq i}\psi\left(\frac{X_j-X_i}{a}\right)Y_j\right)Y_iK'^2\left(\frac{Y_i-y}{h}\right)\\
&:= Q_{11}+Q_{12}.
\end{aligned}
$$

Starting with Q_{11}, we have

$$
\begin{aligned}
\frac{1}{Nh^4}E_\xi(Q_{11}) &= \frac{1}{Nh^4}\frac{1}{a}\psi(0)\int_{\mathbb{R}} y_i^2 K'^2\left(\frac{y_i-y}{h}\right)f(y_i)dy_i\\
&= \frac{1}{Nh^3}\frac{1}{a}\psi(0)\int_{\mathbb{R}}(y+hz)^2K'^2(z)f(y+hz)dz\\
&= \frac{1}{a}\psi(0)\left[\frac{1}{Nh^3}y^2f(y)d_{K'}+\frac{1}{Nh}c^*_{K'}\left\{f(y)+2yf'(y)\right\}+o\left(\frac{1}{Nh}\right)\right].
\end{aligned}
$$

For Q_{12}, notice that

$$
\begin{aligned}
\frac{1}{Nh^4}E_\xi(Q_{12}) &= \frac{1}{Nah^4}\sum_{j\in U, j\neq i}E_\xi\left[\psi\left(\frac{X_j-X_i}{a}\right)Y_iY_jK'^2\left(\frac{Y_i-y}{h}\right)\right]\\
&\overset{iid}{=} \frac{(N-1)}{Nah^4}E_\xi\left[\psi\left(\frac{X_j-X_i}{a}\right)Y_iY_jK'^2\left(\frac{Y_i-y}{h}\right)\right].
\end{aligned} \tag{3.20}
$$

Rewrite model (3.1) as follows:

$$
Y_i = m(X_i)+u_i,\;\; i=1,2,\ldots,N, \tag{3.21}
$$

where $E_\xi(u_i|X_i=x_i)=0$ and $E_\xi(u_i^2|X_i=x_i)=\sigma^2(x_i)$. Using (3.21) in (3.20), we get

$$
\begin{aligned}
&\frac{1}{Nh^4}E_\xi(Q_{12})\\
&\quad= \frac{(N-1)}{Nah^4}E_\xi\left[\{m(X_i)+u_i\}\{m(X_j)+u_j\}\psi\left(\frac{X_j-X_i}{a}\right)K'^2\left(\frac{Y_i-y}{h}\right)\right]\\
&\quad= \frac{(N-1)}{Nah^4}E_\xi\Bigg[\{m(X_i)m(X_j)+m(X_i)u_j+m(X_j)u_i+u_iu_j\}\times\\
&\qquad\qquad\qquad\qquad\psi\left(\frac{X_j-X_i}{a}\right)K'^2\left(\frac{Y_i-y}{h}\right)\Bigg]\\
&\quad:= Q_{121}+Q_{122}+Q_{123}+Q_{124}.
\end{aligned}
$$

$$
\begin{aligned}
&\frac{(N-1)}{Nah^4}E_\xi(Q_{121})\\
&=\frac{(N-1)}{Nah^4}\iiint_{\mathbb{R}} m(x_1)m(x_2)\psi\left(\frac{x_2-x_1}{a}\right)K'^2\left(\frac{y_1-y}{h}\right)t(x_1,y_1)g(x_2)dx_1dx_2dy_1\\
&=\frac{(N-1)}{Nh^3}\iiint_{\mathbb{R}} m(w_1)m(w_1+aw_2)\psi(w_2)K'^2(z)t(w_1,y+hz)g(w_1+aw_2)dw_1dw_2dz\\
&=\frac{(N-1)}{Nh^3}\iiint_{\mathbb{R}} \psi(w_2)K'^2(z)m(w_1)\{m(w_1)+aw_2m'(w_1)+O(a^2)\}\{t(w_1,y)\\
&\qquad\qquad +hzt_2'(w_1,y)+O(h^2)\}\{g(w_1)+aw_2g'(w_1)+O(a^2)\}dw_1dw_2dz\\
&=\frac{(N-1)}{Nh^3}\left[d_{K'}\int_{\mathbb{R}} m^2(w_1)g(w_1)t(w_1,y)dw_1+O(h^2)\right].
\end{aligned}
$$

$$
\begin{aligned}
&\frac{(N-1)}{Nah^4}E_\xi(Q_{122})\\
&=\frac{(N-1)}{Nah^4}\iiiint_{\mathbb{R}} m(x_1)u_2\psi\left(\frac{x_2-x_1}{a}\right)K'^2\left(\frac{m(x_1)+u_1-y}{h}\right)\varphi(x_1,u_1)\varphi(x_2,u_2)\\
&\qquad\qquad dx_1dx_2du_1du_2\\
&=\frac{(N-1)}{Nh^3}\iiiint_{\mathbb{R}} \psi(w_2)K'^2(z_1)m(w_1)z_2\varphi(w_1,y-m(w_1)+hz_1)\varphi(w_1+aw_2,z_2)\\
&\qquad\qquad dw_1dw_2dz_1dz_2\\
&=\frac{(N-1)}{Nh^3}\iiiint_{\mathbb{R}} \psi(w_2)K'^2(z)m(w_1)z_2\{\varphi(w_1,y-m(w_1))+hz_1\varphi_2'(w_1,y-m(w_1))\\
&\qquad\qquad +O(h^2)\}\{\varphi(w_1,z_2)+aw_2\varphi_1'(w_1,z_2)+O(a^2)\}dw_1dw_2dz_1dz_2\\
&=\frac{(N-1)}{Nh^3}\left[d_{K'}\iint_{\mathbb{R}} m(w_1)z_2\varphi(w_1,y-m(w_1))\varphi(w_1,z_2)dw_1dz_2+O(h^2)\right]\\
&=\frac{(N-1)}{Nh^3}\left[d_{K'}\{\int_{\mathbb{R}} m(w_1)g(w_1)\varphi(w_1,y-m(w_1))dw_1\}\{\int_{\mathbb{R}} z_2\varphi(z_2)dz_2\}+O(h^2)\right]\\
&=\frac{(N-1)}{Nh^3}O(h^2),
\end{aligned}
$$

where the last equality follows because $E_\xi(Z_2)=E_\xi(u_2)=0$ and the equality before the last comes from the fact that $\varphi(w_1,z_2)=g(w_1)\varphi(z_2)$ since $w_1=X_1$ and $z_2=u_2$ are independent. Similarly,

$$
\frac{(N-1)}{Nah^4}E_\xi(Q_{123})=\frac{(N-1)}{Nh^3}O(h^2).
$$

Additionally,

$$\begin{aligned}
&\frac{(N-1)}{Nah^4}E_\xi(Q_{124})\\
&=\frac{(N-1)}{Nah^4}\iiiint_{\mathbb{R}} u_1u_2\psi\left(\frac{x_2-x_1}{a}\right)K'^2\left(\frac{m(x_1)+u_1-y}{h}\right)\varphi(x_1,u_1)\varphi(x_2,u_2)\\
&\qquad dx_1dx_2du_1du_2\\
&=\frac{(N-1)}{Nh^3}\iiiint_{\mathbb{R}}\psi(w_2)K'^2(z_1)\{y-m(w_1)+hz_1\}z_2\varphi(w_1+aw_2,z_2)\times\\
&\qquad \varphi(w_1,y-m(w_1)+hz_1)dw_1dw_2dz_1dz_2\\
&=\frac{(N-1)}{Nh^3}\iiiint_{\mathbb{R}}\psi(w_2)K'^2(z)\{y-m(w_1)+hz_1\}z_2\{\varphi(w_1,z_2)+aw_2\varphi_1'(w_1,z_2)\\
&\qquad +O(a^2)\}+aw_2\varphi_1'(w_1,z_2)+O(a^2)\}\{\varphi(w_1,y-m(w_1))\\
&\qquad +hz_1\varphi_2'(w_1,y-m(w_1))+O(h^2)\}dw_1dw_2dz_1dz_2\\
&=\frac{(N-1)}{Nh^3}\left[d_{K'}\iint_{\mathbb{R}}\{y-m(w_1)\}z_2\varphi(w_1,z_2)\varphi(w_1,y-m(w_1))dw_1dz_2+O(h^2)\right]\\
&=\frac{(N-1)}{Nh^3}\left[d_{K'}\int_{\mathbb{R}}\{y-m(w_1)\}g(w_1)\varphi(w_1,y-m(w_1))dw_1\int_{\mathbb{R}}z_2\varphi(z_2)dz_2+O(h^2)\right]\\
&=\frac{(N-1)}{Nh^3}O(h^2).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
E_\xi(Q_2)&=E_\xi\left\{\sum_{j\in U}\psi_a(X_j-X_i)\right\}\\
&=E_\xi\left\{\frac{1}{a}\psi(0)+\sum_{j\in U,j\neq i}\psi_a(X_j-X_i)\right\}\\
&\overset{iid}{=}\frac{1}{a}\psi(0)+(N-1)\iint_{\mathbb{R}}\frac{1}{a}\psi\left(\frac{x_2-x_1}{a}\right)g(x_1)g(x_2)dx_1dx_2\\
&=\frac{1}{a}\psi(0)+(N-1)\iint_{\mathbb{R}}\psi(w_2)g(w_1)g(w_1+aw_2)dw_1dw_2\\
&\approx\frac{1}{a}\psi(0)+(N-1)\iint_{\mathbb{R}}\psi(w_2)g(w_1)\{g(w_1)+aw_2g'(w_1)\}dw_1dw_2\\
&=\frac{1}{a}\psi(0)+(N-1)\int_{\mathbb{R}}g^2(w_1)dw_1\\
&=(N-1)\left[\int_{\mathbb{R}}g^2(w_1)dw_1+\frac{1}{(N-1)a}\psi(0)\right].
\end{aligned}$$

Substituting all results under (3.19) into (3.19), we get

$$\frac{1}{N}E_{\xi}(M_2)=\frac{1}{Nh^3}d_{K'}\{d_g\}^{-1}\int_{\mathbb{R}}m^2(x)g(x)t(x,y)dx+o\left(\frac{1}{Nh^3}\right). \tag{3.22}$$

We now consider M_1. First, observe that

$$\begin{aligned}
\frac{1}{N}E_{\xi}(M_1) &= \frac{1}{Nh^4}E_{\xi}\left\{m_U^2(X_i)K'^2\left(\frac{Y_i-y}{h}\right)\right\}\\
&= \frac{1}{Nh^4}E_{\xi}\left\{\left(\frac{\sum_{j\in U}\psi_a(X_j-X_i)Y_j}{\sum_{j\in U}\psi_a(X_j-X_i)}\right)^2K'^2\left(\frac{Y_i-y}{h}\right)\right\}\\
&= \frac{1}{Nh^4}E_{\xi}\left\{\frac{\left(\sum_{j\in U}\psi_a(X_j-X_i)Y_j\right)^2K'^2\left(\frac{Y_i-y}{h}\right)}{\left(\sum_{j\in U}\psi_a(X_j-X_i)\right)^2}\right\}\\
&:= \frac{1}{Nh^4}E_{\xi}\left\{\frac{Q_3}{Q_4}\right\}\approx\frac{1}{Nh^4}\frac{E_{\xi}(Q_3)}{E_{\xi}(Q_4)}.
\end{aligned}\tag{3.23}$$

Second, expand the squared sum in Q_3 to get

$$\begin{aligned}
Q_3 &= \left(\sum_{j\in U}\psi_a^2(X_j-X_i)Y_j^2+\sum\sum_{j,k\in U,j\neq k}\psi_a(X_j-X_i)\psi_a(X_k-X_i)Y_jY_k\right)K'^2\left(\frac{Y_i-y}{h}\right)\\
&= \frac{1}{a^2}\psi^2(0)Y_i^2K'^2\left(\frac{Y_i-y}{h}\right)+\left(\sum_{j\in U,j\neq i}\psi_a^2(X_j-X_i)Y_j^2\right)K'^2\left(\frac{Y_i-y}{h}\right)\\
&\quad+2\frac{1}{a}\psi(0)Y_iK'^2\left(\frac{Y_i-y}{h}\right)\sum_{k\in U,\ k\neq i}\psi_a(X_k-X_i)Y_k\\
&\quad+\left(\sum\sum_{j,k\in U,j\neq k\neq i}\psi_a(X_j-X_i)\psi_a(X_k-X_i)Y_jY_k\right)K'^2\left(\frac{Y_i-y}{h}\right)\\
&= Q_{31}+Q_{32}+Q_{33}+Q_{34}.
\end{aligned}$$

Next, we evaluate the expectation of each of the four terms, in order, as follows.

$$\begin{aligned}
\frac{1}{Nh^4}E_{\xi}(Q_{31}) &= \frac{1}{Nh^4a^2}\psi^2(0)\int_{\mathbb{R}}y_i^2K'^2\left(\frac{y_i-y}{h}\right)f(y_i)dy_i\\
&= \frac{1}{Nh^3a^2}\psi^2(0)\int_{\mathbb{R}}(y+hz)^2K'^2(z)f(y+hz)dz\\
&= \frac{1}{a^2}\psi^2(0)\left[\frac{1}{Nh^3}y^2f(y)d_{K'}+\frac{1}{Nh}\left\{f(y)+2yf'(y)\right\}c_{K'}^*+o\left(\frac{1}{Nh}\right)\right].
\end{aligned}$$

For Q_{32}, notice that

$$
\begin{aligned}
&\frac{1}{Nh^4}E_\xi(Q_{32})\\
&=\frac{1}{Nh^4}\sum_{j\in U,j\neq i}\iiiint_{\mathbb{R}} y_j^2\psi_a^2(x_j-x_i)K'^2\left(\frac{y_i-y}{h}\right)t(x_i,y_i)t(x_j,y_j)dx_idx_jdy_idy_j\\
&\overset{iid}{=}\frac{(N-1)}{Nh^4a^2}\iiiint_{\mathbb{R}} y_2^2\psi^2\left(\frac{x_2-x_1}{a}\right)K'^2\left(\frac{y_1-y}{h}\right)t(x_1,y_1)t(x_2,y_2)dx_1dx_2dy_1dy_2\\
&=\frac{(N-1)}{Nh^3a}\iiiint_{\mathbb{R}} z_2^2\psi^2(w_2)K'^2(z_1)t(w_1,y+hz_1)t(w_1+aw_2,z_2)dw_1dw_2dz_1dz_2\\
&=\frac{(N-1)}{Nh^3a}\iiiint_{\mathbb{R}} z_2^2\psi^2(w_2)K'^2(z_1)\{t(w_1,y)+hz_1t_2'(w_1,y)+O(h^2)\}\\
&\qquad\qquad\{t(w_1,z_2)+aw_2t_1'(w_1,z_2)+O(a^2)\}dw_1dw_2dz_1dz_2\\
&=\frac{(N-1)}{Nh^3a}\left[d_\psi d_{K'}\iint_{\mathbb{R}} z_2^2t(w_1,y)t(w_1,z_2)dw_1dz_2+O(h^2)\right]\\
&=\frac{(N-1)}{Nh^3a}\left[\mu_{Y^2}d_\psi d_{K'}\int_{\mathbb{R}} g(w_1)t(w_1,y)dw_1+O(h^2)\right].
\end{aligned}
$$

where the fourth equality follows because $\int_{\mathbb{R}} z_1K'^2(z_1)dz_1=0$ and $\int_{\mathbb{R}} w_2\ \psi^2(w_2)dw_2=0$ and the last equality follows from the fact that $t(w_1,z_2)=f(z_2)g(w_1)$ since $W_1=X_1$ and $Z_2=Y_2$ are independent. Considering Q_{33}, we have

$$
\begin{aligned}
&\frac{1}{Nh^4}E_\xi(Q_{33})\\
&=\frac{2(N-1)}{Nh^4a^2}\psi(0)\iiiint_{\mathbb{R}} y_1y_2\psi\left(\frac{x_2-x_1}{a}\right)K'^2\left(\frac{y_1-y}{h}\right)t(x_1,y_1)t(x_2,y_2)dx_1dx_2dy_1dy_2\\
&=\frac{2(N-1)}{Nh^3a}\psi(0)\iiiint_{\mathbb{R}}(y+hz_1)z_2\psi(w_2)K'^2(z_1)t(w_1,y+hz_1)t(w_1+aw_2,z_2)\\
&\qquad\qquad dw_1dw_2dz_1dz_2\\
&=\frac{2(N-1)}{Nh^3a}\psi(0)\iiiint_{\mathbb{R}}\psi(w_2)K'^2(z_1)(y+hz_1)z_2\{t(w_1,y)+hz_1t_2'(w_1,y)+O(h^2)\}\\
&\qquad\qquad\{t(w_1,z_2)+aw_2t_1'(w_1,z_2)+O(a^2)\}dw_1dw_2dz_1dz_2\\
&=\frac{2(N-1)}{Nh^3a}\psi(0)\left[d_{K'}\iint_{\mathbb{R}} yz_2t(w_1,y)t(w_1,z_2)dw_1dz_2+O(h^2)\right]\\
&=\frac{2(N-1)}{Nh^3a}\psi(0)\left[y\mu_Y d_{K'}\int_{\mathbb{R}} g(w_1)t(w_1,y)dw_1+O(h^2)\right].
\end{aligned}
$$

For Q_{34}, using model (3.21), we can write

$$\begin{aligned}
E_{\xi}(Q_{34}) &= E_{\xi}\left\{\sum_{j,k\in U, j\neq k\neq i}\sum \psi_a(X_j-X_i)\psi_a(X_k-X_i)Y_jY_kK'^2\left(\frac{Y_i-y}{h}\right)\right\}\\
&\overset{iid}{=} (N-1)(N-2)E_{\xi}\left\{\psi_a(X_j-X_i)\psi_a(X_k-X_i)Y_jY_kK'^2\left(\frac{Y_i-y}{h}\right)\right\}\\
&= (N-1)(N-2)E_{\xi}\left\{\psi_a(X_j-X_i)\psi_a(X_k-X_i)K'^2\left(\frac{Y_i-y}{h}\right)\times\right.\\
&\qquad\qquad \left.\{m(X_j)+u_j\}\{m(X_k)+u_k\}\right\}\\
&= (N-1)(N-2)E_{\xi}\left\{\psi_a(X_j-X_i)\psi_a(X_k-X_i)K'^2\left(\frac{Y_i-y}{h}\right)\times\right.\\
&\qquad\qquad \left.\{m(X_j)m(X_k)+m(X_j)u_k+u_jm(X_k)+u_ju_k\}\right\}\\
&:= (N-1)(N-2)E_{\xi}\{Q_{341}+Q_{342}+Q_{344}+Q_{344}\}.
\end{aligned} \tag{3.24}$$

Then,

$$\begin{aligned}
&E_{\xi}(Q_{341})\\
&= \iiiint_{\mathbb{R}} \frac{1}{a^2}m(x_2)m(x_3)\psi\left(\frac{x_2-x_1}{a}\right)\psi\left(\frac{x_3-x_1}{a}\right)K'^2\left(\frac{y_1-y}{h}\right)t(x_1,y_1)g(x_2)g(x_3)\\
&\qquad dx_1dx_2dx_3dy_1\\
&= h\iiiint_{\mathbb{R}} m(w_1+aw_2)m(w_1+aw_3)\psi(w_2)\psi(w_3)K'^2(z)t(w_1,y+hz)g(w_1+aw_2)\times\\
&\qquad g(w_1+aw_3)dw_1dw_2dw_3dz\\
&= h\iiiint_{\mathbb{R}} \psi(w_2)\psi(w_3)K'^2(z)\{m(w_1)+aw_2m'(w_1)+O(a^2)\}\{m(w_1)+aw_3m'(w_1)\\
&\qquad +O(a^2)\}\{t(w_1,y)+hzt_2'(w_1,y)+O(h^2)\}\{g(w_1)+aw_2g'(w_1)\\
&\qquad +O(a^2)\}\{g(w_1)+aw_3g'(w_1)+O(a^2)\}dw_1dw_2dw_3dz\\
&= h\left[d_{K'}\int m^2(w_1)g^2(w_1)t(w_1,y)dw_1+O(h^2)\right].
\end{aligned}$$

$$\begin{aligned}
&E_{\xi}(Q_{342})\\
&= \iiiint_{\mathbb{R}} \frac{1}{a^2}m(x_2)u_3\psi\left(\frac{x_2-x_1}{a}\right)\psi\left(\frac{x_3-x_1}{a}\right)K'^2\left(\frac{m(x_1)+u_1-y}{h}\right)\varphi(x_1,u_1)\times\\
&\qquad g(x_2)\varphi(x_3,u_3)dx_1dx_2dx_3du_1du_3
\end{aligned}$$

$$
\begin{aligned}
&= h\iiiint_{\mathbb{R}} m(w_1+aw_2)z_3\psi(w_2)\psi(w_3)K'^2(z_1)\varphi(w_1,y-m(w_1)+hz_1)\times \\
&\qquad\qquad g(w_1+aw_2)\varphi(w_1+aw_3,z_3)dw_1dw_2dw_3dz_1dz_3 \\
&= h\iiiint_{\mathbb{R}} z_3\psi(w_2)\psi(w_3)K'^2(z_1)\{m(w_1)+aw_2m'(w_1)+O(a^2)\}\{\varphi(w_1,y-m(w_1)) \\
&\qquad +hz_1\varphi_2'(w_1,y-m(w_1))+O(h^2)\}\{g(w_1)+aw_2g'(w_1)+O(a^2)\}\times \\
&\qquad \{\varphi(w_1,z_3)+aw_3\varphi_1'(w_1,z_3)+O(a^2)\}dw_1dw_2dw_3dz_1dz_3 \\
&= h\left[d_{K'}\iint z_3m(w_1)\varphi(w_1,y-m(w_1))g(w_1)\varphi(w_1,z_3)dw_1dz_3+O(h^2)\right] \\
&= h\left[d_{K'}\{\int m(w_1)g^2(w_1)\varphi(w_1,y-m(w_1))dw_1\}\{\int z_3\varphi(z_3)dz_3\}+O(h^2)\right] \\
&= O(h^3).
\end{aligned}
$$

Similarly, $E_\xi(Q_{343}) = O(h^3)$ and $E_\xi(Q_{344}) = O(h^3)$. Using these resuts in (3.24), we have

$$
\frac{1}{Nh^4}E_\xi(Q_{34}) = \frac{1}{Nh^3}\left[d_{K'}\int m^2(w_1)g^2(w_1)t(w_1,y)dw_1+O(h^2)\right]
$$

On the other hand,

$$
\begin{aligned}
E_\xi(Q_4) &= E_\xi\left\{\sum_{j\in U}\psi_a^2(X_j-X_i)+\sum\sum_{j,k\in U,j\neq k}\psi_a(X_j-X_i)\psi_a(X_k-X_i)\right\} \\
&= E_\xi(Q_{41})+E_\xi(Q_{42}).
\end{aligned}
$$

Evaluate each of the two terms separately. First, observe that

$$
\begin{aligned}
E_\xi(Q_{41}) &\overset{iid}{=} \frac{1}{a^2}\psi^2(0)+\frac{(N-1)}{a^2}\iint_{\mathbb{R}}\psi^2\left(\frac{x_2-x_1}{a}\right)g(x_1)g(x_2)dx_1dx_2 \\
&= \frac{1}{a^2}\psi^2(0)+\frac{(N-1)}{a}\iint_{\mathbb{R}}\psi^2(w_2)g(w_1)g(w_1+aw_2)dw_1dw_2 \\
&\approx \frac{1}{a^2}\psi^2(0)+\frac{(N-1)}{a}\iint_{\mathbb{R}}\psi^2(w_2)g(w_1)\left\{g(w_1)+aw_2g'(w_1)\right\}dw_1dw_2 \\
&= \frac{1}{a^2}\psi^2(0)+(N-1)\left[\frac{1}{a}d_\psi d_g+O(a)\right].
\end{aligned}
$$

Second,

$$\begin{aligned}Q_{42} &= 2\frac{1}{a}\psi(0)\sum_{k\in U,\ k\neq i}\psi_a(X_k-X_i)+\sum_{j,k\in U,j\neq k\neq i}\sum\psi_a(X_j-X_i)\psi_a(X_k-X_i)\\ &:= Q_{421}+Q_{422}.\end{aligned}$$

$$\begin{aligned}E_\xi(Q_{421}) &= 2\frac{1}{a}\psi(0)\sum_{k\in U,\ k\neq i}\iint_{\mathbb{R}}\psi_a(x_k-x_i)g(x_i)g(x_k)dx_idx_k\\ &= 2\frac{1}{a}\psi(0)(N-1)\iint_{\mathbb{R}}\psi(w_2)g(w_1)g(w_1+aw_2)dw_1dw_2\\ &= 2\frac{1}{a}\psi(0)(N-1)\left[\int_{\mathbb{R}}g^2(w_1)dw_1+O(a^2)\right].\end{aligned}$$

$$\begin{aligned}&E_\xi(Q_{422})\\ &= \sum_{j,k\in U,j\neq k\neq i}\sum\iiint_{\mathbb{R}}\psi_a(x_j-x_i)\psi_a(x_k-x_i)g(x_i)g(x_j)g(x_k)dx_idx_jdx_k\\ &= (N-1)(N-2)\iiint_{\mathbb{R}}\psi(w_2)\psi(w_3)g(w_1)g(w_1+aw_2)g(w_1+aw_3)dw_1dw_2dw_3\\ &= (N-1)(N-2)\left[\int_{\mathbb{R}}g^3(w_1)dw_1+O(a^2)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}E_\xi(Q_4) &= (N-1)(N-2)\left[\int_{\mathbb{R}}g^3(w_1)dw_1+\frac{1}{(N-2)a}\left\{2\psi(0)+d_\psi\right\}d_g+O(a^2)\right]\\ &= (N-1)(N-2)\left[\int_{\mathbb{R}}g^3(w_1)dw_1+O\left(\frac{1}{Na}+a^2\right)\right].\end{aligned}$$

Letting $S_g=\int_{\mathbb{R}}g^3(x)dx$ and substituting all results after (3.23) into (3.23), we get

$$\frac{1}{N}E_\xi(M_1)=\frac{1}{Nh^3}d_{K'}\{S_g\}^{-1}\int m^2(x)g(x)t(x,y)dx+o\left(\frac{1}{Nh^3}\right). \tag{3.25}$$

Now, we use (3.18), (3.22) and (3.25) in (3.17) to get

$$\frac{1}{N}E_\xi(W_i^2)=\frac{1}{Nh^3}d_{K'}f(y)\left[\{S(g)\}^{-1}\int m^2(x)g^2(x)t(x|y)dx\right.$$

$$-2\{d_g\}^{-1}\int_{\mathbb{R}} m^2(x)g(x)t(x|y)dx+y^2\Bigg]+o\left(\frac{1}{Nh^3}\right). \quad (3.26)$$

To evaluate the second term in (3.16), notice that by a Taylor expansion for $K_h(y-\hat{Y}_{Ul})$ about Y_l and some algebra similar to those in (2.34) in the proof of Theorem 2.2, we have

$$\begin{aligned} W_iW_j &= \left\{K_h(y-Y_i)-K_h(y-\hat{Y}_{Ui})\right\}\left\{K_h(y-Y_j)-K_h(y-\hat{Y}_{Uj})\right\} \\ &\approx \frac{1}{h^4}(\hat{Y}_{Ui}\hat{Y}_{Uj}-Y_i\hat{Y}_{Uj}+Y_iY_j)K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\ &:= \chi_1-2\chi_2+\chi_3. \end{aligned} \quad (3.27)$$

We now examine the expectation of each of the three terms in (3.27) in reverse. First,

$$\begin{aligned} \frac{1}{n}E_\xi(\chi_3) &= \frac{1}{nh^4}E_\xi\left[Y_iY_jK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\ &\overset{iid}{=} \frac{1}{nh^4}\left\{E_\xi\left[Y_1K'\left(\frac{Y_1-y}{h}\right)\right]\right\}^2 \\ &= \frac{1}{nh^4}\left\{\int_{\mathbb{R}} y_1K'\left(\frac{y_1-y}{h}\right)f(y_1)dy_1\right\}^2 \\ &= \frac{1}{nh^4}\left\{h\int_{\mathbb{R}}(y+hz)K'(z)f(y+hz)dz\right\}^2 \\ &= \frac{1}{nh^2}\left\{yf(y)\int_{\mathbb{R}}K'(z)dz+h[f(y)+yf'(y)]\int_{\mathbb{R}}zK'(z)dz+\cdots\right\}^2 \\ &= \frac{1}{n}\left\{[f(y)+yf'(y)]\int_{\mathbb{R}}zK'(z)dz+\cdots\right\}^2, \end{aligned} \quad (3.28)$$

where (3.28) follows from the fact that $\int_{\mathbb{R}} K'(z)dz=0$ by the assumptions on K. Next, we consider χ_2.

$$\begin{aligned} \frac{1}{n}E_\xi(\chi_2) &= \frac{1}{nh^4}E_\xi\left[\hat{Y}_{Uj}Y_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\ &= \frac{1}{nh^4}E_\xi\left[\left(\frac{\sum_{k\in U}\psi_a(X_k-X_j)Y_k}{\sum_{k\in U}\psi_a(X_k-X_j)}\right)Y_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\ &= \frac{1}{nh^4}E_\xi\left[\frac{\left(\sum_{k\in U}\psi\left(\frac{X_k-X_j}{a}\right)Y_k\right)Y_iK'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)}{\sum_{k\in U}\psi\left(\frac{X_k-X_j}{a}\right)}\right] \\ &:= \frac{1}{nh^4}E_\xi\left[\frac{\omega_1}{\omega_2}\right]\approx\frac{1}{nh^4}\frac{E_\xi(\omega_1)}{E_\xi(\omega_2)}. \end{aligned} \quad (3.29)$$

First, observe that

$$\begin{aligned}\omega_1 &= \left[\psi(0)Y_jY_i + \psi\left(\frac{X_i - X_j}{a}\right)Y_i^2 + \sum_{k\in U, k\neq i\neq j}\psi\left(\frac{X_k - X_j}{a}\right)Y_kY_i\right]\\ &\quad \times K'\left(\frac{Y_i - y}{h}\right)K'\left(\frac{Y_j - y}{h}\right)\\ &:= \omega_{11} + \omega_{12} + \omega_{13}.\end{aligned}$$

Second, we evaluate the expectation of the three terms in order. Considering ω_{11}, we have

$$\begin{aligned}\frac{1}{nh^4}E_\xi(\omega_{11}) &\overset{iid}{=} \frac{1}{nh^4}\psi(0)E_\xi\left[Y_1K'\left(\frac{Y_1 - y}{h}\right)\right]E_\xi\left[Y_2K'\left(\frac{Y_2 - y}{h}\right)\right]\\ &= \frac{1}{nh^4}\psi(0)\left[\int_{\mathbb{R}} y_1K'\left(\frac{y_1 - y}{h}\right)f(y_1)dy_1\right]^2\\ &= \frac{1}{nh^2}\psi(0)\left[\int_{\mathbb{R}}(y+hz)K'(z)f(y+hz)dz\right]^2\\ &= \frac{1}{nh^2}\psi(0)\left[h\left\{f(y) + yf'(y)\right\}\int_{\mathbb{R}} zK'(z)dz\right]^2\\ &= O\left(\frac{1}{n}\right).\end{aligned}$$

For ω_{12}, we have

$$\begin{aligned}&\frac{1}{nh^4}E_\xi(\omega_{12})\\ &\overset{iid}{=} \frac{1}{nh^4}E_\xi\left[\psi\left(\frac{X_1 - X_2}{a}\right)Y_1^2K'\left(\frac{Y_1 - y}{h}\right)K'\left(\frac{Y_2 - y}{h}\right)\right]\\ &= \frac{1}{nh^4}\iiiint_{\mathbb{R}} y_1^2\psi\left(\frac{x_1 - x_2}{a}\right)K'\left(\frac{y_1 - y}{h}\right)K'\left(\frac{y_2 - y}{h}\right)t(x_1,y_1)t(x_2,y_2)dx_1dx_2dy_1dy_2\\ &= \frac{a}{nh^2}\iiiint_{\mathbb{R}}(y+hz_1)^2\psi(w_2)K'(z_1)K'(z_2)t(w_1, y+hz_1)t(w_1+aw_2, y+hz_2)\\ &\qquad\qquad dw_1dw_2dz_1dz_2\\ &\approx \frac{a}{nh^2}\iiiint_{\mathbb{R}}(y+hz_1)^2\psi(w_2)K'(z_1)K'(z_2)\left\{t(y|w_1) + hz_1t_1'(y|w_1)\right\}\left\{t(y|w_1)\right.\\ &\qquad \left. + hz_2t_1'(y|w_1) + aw_2t_2'(y|w_1)\right\}g^2(w_1)dw_1dw_2dz_1dz_2\\ &= \frac{a}{nh^2}\int_{\mathbb{R}} h^2\left\{y^2[t_1'(y|w_1)]^2 + 2yt_1'(y|w_1) + O(h^3)\right\}g^2(w_1)dw_1\\ &= o\left(\frac{1}{n}\right).\end{aligned}$$

Consider ω_{13} and observe that

$$
\begin{aligned}
&\frac{1}{nh^4}E_\xi(\omega_{13})\\
&\overset{iid}{=} \frac{1}{nh^4}(N-2)E_\xi\left[\psi\left(\frac{X_3-X_2}{a}\right)Y_3Y_1K'\left(\frac{Y_1-y}{h}\right)K'\left(\frac{Y_2-y}{h}\right)\right]\\
&= \frac{(N-2)}{nh^4}\left[\int_{\mathbb{R}} y_1K'\left(\frac{y_1-y}{h}\right)f(y_1)dy_1\right]\\
&\quad\left[\iiiint_{\mathbb{R}} y_3\psi\left(\frac{x_3-x_2}{a}\right)K'\left(\frac{y_2-y}{h}\right)t(x_2,y_2)t(x_3,y_3)dx_2dx_3dy_2dy_3\right]\\
&= \frac{(N-2)a}{nh^2}\left[\int_{\mathbb{R}}(y+hz_1)K'(z_1)f(y+hz_1)dz_1\right]\left[\iiiint_{\mathbb{R}} z_3(y+hz_2)\psi(w_2)\right.\\
&\qquad\qquad \left.\times K'(z_2)t(w_1,y+hz_2)t(w_1+aw_2,z_3)dw_1dw_2dz_2dz_3\right]\\
&\approx \frac{(N-2)a}{nh^2}\left[h\left\{f(y)+yf'(y)\right\}\int_{\mathbb{R}} z_1K'(z_1)\,dz_1\right]\\
&\quad\left[h\int_{\mathbb{R}} z_3\left\{yt_1'(y|w_1)t(w_1,z_3)+t(y|w_1)t(w_1,z_3)\right\}g(w_1)dw_1dz_3\int_{\mathbb{R}} z_2K'(z_2)\,dz_2\right]\\
&= \frac{(N-2)a}{n}\left\{f(y)+yf'(y)\right\}\int_{\mathbb{R}} z\left\{yt_1'(y|w)t(w,z)+t(y|w)t(w,z)\right\}g(w)dwdz.
\end{aligned}
$$

For the denominator of (3.29), notice that

$$
\begin{aligned}
E_\xi(\omega_2) &= E_\xi\left[\psi(0)+\sum_{i\in U,i\neq j}\psi\left(\frac{X_j-X_i}{a}\right)\right]\\
&\overset{iid}{=} \psi(0)+(N-1)\iint_{\mathbb{R}}\psi\left(\frac{x_2-x_1}{a}\right)g(x_1)g(x_2)dx_1dx_2\\
&= \psi(0)+(N-1)a\iint_{\mathbb{R}}\psi(w_2)g(w_1)g(w_1+aw_2)dw_1dw_2\\
&= \psi(0)+(N-1)a\times\\
&\quad \iint_{\mathbb{R}}\psi(w_2)g(w_1)\{g(w_1)+aw_2g'(w_1)+\frac{1}{2}a^2w_2^2g''(w_1)+o(a^2)\}dw_1dw_2\\
&= \psi(0)+(N-1)a\{\int_{\mathbb{R}}g^2(w)dw+c_\psi\int_{\mathbb{R}}g''(w)g(w)dw+o(a^2)\}.
\end{aligned}
$$

Substituting these results into (3.29), we get

$$
\begin{aligned}
\frac{1}{n}E_\xi(\chi_2) = \frac{(N-2)a}{n(N-1)a}&\left\{f(y)+yf'(y)\right\}\int_{\mathbb{R}} z\left\{yt_1'(y|w)t(w,z)+t(y|w)t(w,z)\right\}g(w)dwdz\\
&\left[\int_{\mathbb{R}}g^2(w)dw+c_\psi\int_{\mathbb{R}}g''(w)g(w)dw+o(a^2)\right]^{-1}+O\left(\frac{1}{nNa}\right)
\end{aligned}
$$

$$= O\left(\frac{1}{n}\right). \tag{3.30}$$

Next, we work on χ_1. First, observe that

$$\begin{aligned}
&\frac{1}{n}E_\xi(\chi_1) \\
&= \frac{1}{nh^4}E_\xi\left[\left(\frac{\sum_{k\in U}\psi_a(X_k-X_i)Y_k}{\sum_{k\in U}\psi_a(X_k-X_i)}\right)\left(\frac{\sum_{k\in U}\psi_a(X_k-X_j)Y_k}{\sum_{k\in U}\psi_a(X_k-X_j)}\right)K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)\right] \\
&= \frac{1}{nh^4}E_\xi\left[\frac{\left(\sum_{k\in U}\psi_a(X_k-X_i)Y_k\right)\left(\sum_{k\in U}\psi_a(X_k-X_j)Y_k\right)K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right)}{\left(\sum_{k\in U}\psi_a(X_k-X_i)Y_k\right)\left(\sum_{k\in U}\psi_a(X_k-X_i)Y_k\right)}\right] \\
&:= \frac{1}{nh^4}E_\xi\left[\frac{\omega_3}{\omega_4}\right] \approx \frac{1}{nh^4}\frac{E_\xi(\omega_3)}{E_\xi(\omega_4)}. \qquad (3.31)
\end{aligned}$$

But,

$$\begin{aligned}
\omega_3 &= \left(\sum_{k\in U}\psi_a(X_k-X_i)\psi_a(X_k-X_j)Y_k^2 + \sum\sum_{k,l\in U,\ k\neq l}\psi_a(X_k-X_i)Y_k\psi_a(X_l-X_j)Y_l\right) \\
&\qquad \times K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\
&= \Bigg[2\psi_a(0)\psi_a(X_j-X_i)Y_j^2 + \sum_{k\in U,\ k\neq i\neq j}\psi_a(X_k-X_i)\psi_a(X_k-X_j)Y_k^2 + 2\psi_a^2(0)Y_iY_j \\
&\quad +2\sum_{k\in U,\ k\neq i\neq j}\psi_a(X_k-X_i)\psi_a(X_i-X_j)Y_iY_k \\
&\quad + \sum\sum_{k,l\in U,\ k\neq l\neq i\neq j}\psi_a(X_k-X_i)\psi_a(X_k-X_j)Y_kY_l\Bigg]K'\left(\frac{Y_i-y}{h}\right)K'\left(\frac{Y_j-y}{h}\right) \\
&:= \omega_{31}+\omega_{32}+\omega_{33}+\omega_{34}+\omega_{35}.
\end{aligned}$$

Using manipulations similar to the ones we used in obtaining 3.28 and 3.30, we can evaluate the expectation of each of the five terms and show that

$$\frac{1}{nh^4}E_\xi(\omega_3) = O\left(\frac{(N-2)(N-3)}{n}\right).$$

On the other hand, similar manipulations show that

$$E_\xi(\omega_4)$$

$$
\begin{aligned}
&= E_\xi\left(\sum_{k\in U}\psi_a(X_k-X_i)\psi_a(X_k-X_j)Y_k^2+\sum\sum_{k,l\in U,\ k\neq l}\psi_a(X_k-X_i)Y_k\psi_a(X_l-X_j)Y_l\right)\\
&= \sum_{k\in U}E_\xi\{\psi_a(X_k-X_i)\psi_a(X_k-X_j)Y_k^2\}+\sum\sum_{k\in U,\ k\neq l}E_\xi\{\psi_a(X_k-X_i)Y_k\psi_a(X_l-X_j)Y_l\}\\
&= O\{N(N-1)\}.
\end{aligned}
$$

Substituting these results into (3.31), we get

$$
\frac{1}{n}E_\xi(\chi_1)=O\left(\frac{1}{n}\right). \tag{3.32}
$$

Using (3.28), (3.30) and (3.32) in (3.27), we get

$$
\frac{1}{n}E_\xi(W_iW_j)=O\left(\frac{1}{n}\right). \tag{3.33}
$$

Now, substitute (3.26) and (3.33) into (3.16) to get

$$
\begin{aligned}
L_1=\frac{1}{Nh^3}d_{K'}f(y)\Bigg[&\{S_g\}^{-1}\int m^2(x)g^2(x)t(x|y)dx\\
&-2\{d_g\}^{-1}\int_{\mathbb{R}}m^2(x)g(x)t(x|y)dx+y^2\Bigg]+o\left(\frac{1}{Nh^3}\right).
\end{aligned} \tag{3.34}
$$

Finally, using (3.13) and (3.34) in (3.12), we get

$$
\begin{aligned}
V_C\left[\hat{f}_{dn}(y;h)\right]=\frac{1}{Nh^3}\left(\frac{1}{N}\sum_{i\in U}\Delta_i\right)d_{K'}f(y)\Bigg[&\{S_g\}^{-1}\int m^2(x)g^2(x)t(x|y)dx\\
&-2\{d_g\}^{-1}\int_{\mathbb{R}}m^2(x)g(x)t(x|y)dx+y^2\Bigg]+o\left(\frac{1}{Nh^3}\right).
\end{aligned} \tag{3.35}
$$

Integrating the variance in (3.35) over y, we obtain the integrated variance of $\hat{f}_{dn}(\cdot;h)$:

$$
\begin{aligned}
IV_C\left[\hat{f}_{dn}(\cdot;h)\right]=\frac{1}{Nh^3}\left(\frac{1}{N}\sum_{i\in U}\Delta_i\right)d_{K'}\Big[&\{S_g\}^{-1}E_\xi\{m^2(X)g^2(X)\}\\
&-2\{d_g\}^{-1}E_\xi\{m^2(X)g(X)\}+\mu_{Y^2}\Big]+o\left(\frac{1}{Nh^3}\right).
\end{aligned} \tag{3.36}
$$

The integrated squared bias of $\hat{f}_{dn}(\cdot;h)$ is obtained by integrating (3.8) over y:

$$ISB_C\left[\hat{f}_{dn}(\cdot;h)\right] = \frac{1}{4}h^4c_K^2\int_{\mathbb{R}}\{f''(y)\}^2dy + o(h^4). \tag{3.37}$$

Adding (3.36) to (3.37) gives the MISE of $\hat{f}_{dn}(\cdot;h)$ and the proof is complete. □

3.4. Asymptotic Distribution of $\hat{f}_{dn}(y)$

In this section, the asymptotic distribution of the estimator $\hat{f}_{dn}(y;h)$ is derived under both the design-based and the combined inference frameworks. As indicated in Section 2.4, our analysis will be restricted to the case where the sampling design used is SRSWOR. We start by the following lemma which gives the asymptotic distribution of a standardized version of $\hat{f}_{dn}(y;h)$ in the design space. This result can be useful for making inference about the finite population smooth $f_U(y;h)$ based on $\hat{f}_{dn}(y;h)$. This lemma is also useful for proving Theorem 3.3 which serves as a central limit theorem for $\hat{f}_{dn}(y;h)$ in the product space.

Lemma 3.2. *Suppose Assumptions A.1–A.4 in Section 2.2.2 hold. Further, suppose that $K(x) \leq M$ for all $x \in \mathbb{R}$ and $\sqrt{n}h \to \infty$ as $n \to \infty$. Then, under SRSWOR, we have*

$$\frac{\hat{f}_{dn}(y;h) - f_U(y;h)}{\hat{\Psi}_{\mathcal{P}}^{1/2}} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1), \tag{3.38}$$

where

$$\hat{\Psi}_{\mathcal{P}} = \left(1 - \frac{n}{N}\right)\frac{\left[\sum_{i\in s}(u_i(h) - z_i(h))^2 - n^{-1}\{\sum_{i\in s}(u_i(h) - z_i(h))\}^2\right]}{n(n-1)},$$

with $u_i(h) = K_h(y - y_i)$ and $z_i(h) = K_h(y - \hat{m}(x_i))$ where $\hat{m}(x_i)$ is as defined in (3.3).

Proof: Follows the same lines of the proof of Lemma 2.3. □

Theorem 3.3. *Suppose Assumptions A.1–A.4 and B.1–B.3 hold. Further, suppose that $K(x) \leq M$ for all $x \in \mathbb{R}$ and $\sqrt{n}h \to \infty$ as $n \to \infty$. Then, if we take $\hat{m}(\cdot)$ in* (3.5) *to be* *the Nadaraya-Watson estimator* $\hat{m}(\cdot;0,a)$ *and use the SRSWOR design, we get*

$$\frac{\hat{f}_{dn}(y;h) - E_\xi\{f_U(y;h)\}}{\sqrt{V_C\{\hat{f}_{dn}(y;h)\}}} \xrightarrow{\mathscr{L}_C} N(0,1), \tag{3.39}$$

where $E_\xi\{f_U(y;h)\}$ *is given by* (2.20) *and*

$$\begin{aligned} V_C\left[\hat{f}_{dn}(y;h)\right] &= \left(1-\frac{n}{N}\right)\frac{1}{nh^3}f(y)d_{K'}\left[\{S_g\}^{-1}\int m^2(x)g^2(x)t(x|y)dx\right. \\ &\qquad \left. -2\{d_g\}^{-1}\int_{\mathbb{R}} m^2(x)g(x)t(x|y)dx + y^2\right] + o\left(\frac{1}{Nh^3}\right) \\ &= \left(1-\frac{n}{N}\right)\frac{1}{nh^3}f(y)d_{K'}\chi(x,y) + o\left(\frac{1}{Nh^3}\right), \end{aligned} \tag{3.40}$$

with $d_g = \int g^2(x)dx$ *and* $S_g = \int g^3(x)dx$.

Proof: Follows the same lines of the proof of Theorem 2.3. □

3.5. Bandwidth Selection

The density estimator $\hat{f}_{dn}(y;h)$, with $\hat{m}(\cdot) = \hat{m}(x;0,a)$ in (3.5), uses two smoothing parameters; the first bandwidth parameter, a, is used to smooth the regression estimate $\hat{m}(x;0,a)$ (see Eq. (3.4)) while the second bandwidth parameter, h, determines the amount of smoothness of the density estimate itself. In this section, we discuss how these parameters can be chosen such that a good balance between the bias and the variance, is achieved.

Starting with the bandwidth a, recall from (3.4) that the design-weighted local constant regression estimator at any point x is given by:

$$\hat{m}(x;0,a) = \frac{\sum_{j\in s} d_j y_j \psi_a(x_j - x)}{\sum_{j\in s} d_j \psi_a(x_j - x)}. \tag{3.41}$$

It can be shown that the sampling design does not affect the bias of local polynomial kernel regression estimators (e.g., Harms and Duchesne (2010) and Johnson et al. (2008)). The design effect on such estimators takes the from of a factor multiplied by the variance of the pure model-based local polynomial regression estimator (cf. Harms and Duchesne (2010)). Given these notes, it can be shown that the MSE of $\hat{m}(x;0,a)$, under the combined inference approach, has the following form:

$$MSE_C\left[\hat{m}(x;0,a)\right] = \frac{1}{na}\frac{\sigma^2(x)}{g(x)}\delta^* d_\psi + \frac{1}{4}a^4[B(x)]^2 c_\psi^2 + o\left(a^4 + \frac{1}{Na}\right), \tag{3.42}$$

where $\delta^* = \left[nN^{-2}\sum_{i\in U}\Delta_i + n/N\right]$ and $B(x) = [m''(x) + \{2m'(x)g'(x)\}/g(x)]$. Integrating the leading terms of the above MSE multiplied by the weighting function $w(x) = g(x)w_0(x)$ for some specific function $w_0(x)$, and taking $\sigma(x) = \sigma$ we get the following form for the asymptotic weighted MISE (AWMISE):

$$AWMISE\left[\hat{m}(\cdot;0,a)\right] = \frac{1}{na}\delta^* d_\psi \sigma^2 \int_{\mathbb{R}} w_0(x)dx + \frac{1}{4}a^4 c_\psi^2 \int_{\mathbb{R}} [B(x)]^2 w_0(x) g(x) dx.$$

Then, the bandwidth that minimizes the above $AWMISE$ is

$$a_{opt} = \left[\frac{\delta^* d_\psi \sigma^2 \int_{\mathbb{R}} w_0(x)dx}{nc_\psi^2 \int_{\mathbb{R}}[B(x)]^2 w_0(x)g(x)dx}\right]^{1/5}.$$

Now, to estimate a_{opt} from the data, one needs to get estimates for σ and for the functions $g(x), g'(x), m'(x)$ and $m''(x)$ which are embedded in $B(x)$. For both functions $g(x)$ and $g'(x)$, one can use kernel density and derivative estimators with normal rule-of-thumb smoothing parameters. On the other hand, we can use global third order parametric polynomial fits to estimate the functions $m(x)$, $m'(x)$ and $m''(x)$ (e.g., Fan and Gijbels (1996, sec. 4.2)). Using these estimates, we can get an estimate for $B(x)$ which we denote by $\tilde{B}(x)$. Finally, an estimate for σ, say $\tilde{\sigma}$, can be obtained from the residuals of the parametric estimate of $m(x)$. Then, a rule-of-thumb estimator of a_{opt} can be obtained as:

$$\hat{a}_{ROT} = \left[\frac{\delta^* d_\psi \tilde{\sigma}^2 \int_{\mathbb{R}} w_0(x)dx}{c_\psi^2 \sum_{i=1}^{n}[\tilde{B}(x_i)]^2 w_0(x_i)}\right]^{1/5}. \tag{3.43}$$

Fan and Gijbels (1996, sec. 4.2) give rule-of-thumb bandwidth estimators of the same form as (3.43) for general local polynomial regression estimators of order p.

Now, we consider selecting the amount of smoothing for the density estimate itself (h). Recall from Theorem 3.2 that the asymptotic MISE of $\hat{f}_{dn}(y;h)$ is given by

$$AMISE\left[\hat{f}_{dn}(\cdot;h)\right] = \frac{1}{nh^3}\delta D d_{K'} + \frac{1}{4}h^4 c_K^2 d_{f''}, \tag{3.44}$$

where

$$D = \left[\frac{E_\xi\{m^2(X)g^2(X)\}}{S_g} - 2\frac{E_\xi\{m^2(X)g(X)\}}{d_g} + \mu_{Y^2}\right], \tag{3.45}$$

$d_g = \int g^2(x)dx$ and $S_g = \int g^3(x)dx$. Under this error criterion, the asymptotically optimal bandwidth $h_{opt,dn}$, in the sense that it minimizes the $AMISE_C$, is obtained using some simple calculus:

$$h_{opt,dn} = \left[\frac{3\delta D d_{K'}}{c_K^2 d_{f''}}\right]^{1/7} n^{-1/7}. \tag{3.46}$$

A direct plug-in estimator of $h_{opt,dn}$ can be obtained by finding estimates for the quantities D and $d_{f''}$ and substituting these estimates into (3.46). Let $I_1 = E_\xi\{m^2(X)g^2(X)\}$ and $I_2 = E_\xi\{m^2(X)g(X)\}$. We use the following quantities to build an estimator for D, say $\tilde{D}$:

$$\hat{\mu}_{Y^2} = \tfrac{1}{n}\sum_{i=1}^n y_i^2, \;\; \hat{d}_g = \tfrac{1}{n}\sum_{i=1}^n \hat{g}(x_i;b_1), \;\; \hat{S}_g = \tfrac{1}{n}\sum_{i=1}^n \hat{g}^2(x_i;b_1),$$

$$\hat{I}_1 = \tfrac{1}{n}\sum_{i=1}^n \hat{m}^2(x_i;a_1)\hat{g}^2(x_i;b_1) \quad \text{and} \quad \hat{I}_2 = \tfrac{1}{n}\sum_{i=1}^n \hat{m}^2(x_i;a_1)\hat{g}(x_i;b_1),$$

where

$$\hat{g}(x;b_1) = \tfrac{1}{nb_1}\textstyle\sum_{i=1}^n K\left(\frac{x-x_i}{b_1}\right) \text{ and } \hat{m}(x;a_1) = \frac{\sum_{i=1}^n y_i K\{a_1^{-1}(x-x_i)\}}{\sum_{i=1}^n K\{a_1^{-1}(x-x_i)\}},$$

with b_1 being a direct plug-in bandwidth obtained following the steps described in Section 2.5 while a_1 is a rule-of-thumb bandwidth similar to the bandwidth in (3.43). Substituting the above quantities into (3.45), we get $\tilde{D}$. Plugging $\tilde{D}$ into (3.46) and using the same four steps listed at the end of Section 2.5, we get the following estimate of $h_{opt,dn}$:

$$\hat{h}_{DPI,dn} = \left[\frac{3\delta \tilde{D} d_{K'}}{n c_K^2 \hat{\Psi}_4(b^*)} \right]^{1/7}. \tag{3.47}$$

The estimators $\hat{a}_{ROT}$ and $\hat{h}_{DPI,dn}$ are used in our simulation study of Chapter 5.

CHAPTER 4

KDE Using Auxiliary Information Via Nonparametric Regression Models of A Kernel-Transformed Study Variable

4.1. Introduction

Recall from Section 2.2.1 that the standard Rosenblatt-Parzen kernel density estimator of $f(y)$ based on the entire finite population can be written as follows:

$$f_U(y;h) = \frac{1}{Nh}\left[\sum_{i\in s} K\left(\frac{y-y_i}{h}\right) + \sum_{i\in\bar{s}} K\left(\frac{y-Y_i}{h}\right)\right]. \tag{4.1}$$

As we noted earlier, the second term on the right-hand side of (4.1) is unknown as it contains the non-sampled Y values; $\{Y_i : i \in \bar{s}\}$. Using the available auxiliary data to predict this term can lead to both model-based and model-assisted density estimators for $f(y)$. In Chapters 2 and 3, both parametric and nonparametric regression models were used to predict the term $\sum_{i\in\bar{s}} K(h^{-1}\{y - Y_i\})$ in (4.1) through replacing the unobserved Y values by the fitted values $\hat{Y}$. Apparently, this approach is a plug-in approach that predicts the unobserved Y values inside the kernel function. Using this plug-in approach, two model-assisted kernel density estimators were proposed for estimating $f(y)$ and the statistical properties of these estimators were investigated in Chapters 2 and 3. It was found that the order of the smoothing parameter for each of the two new estimators is $O(n^{-1/7})$ which is different from the order $O(n^{-1/5})$ commonly obtained for classical kernel density estimators that use no auxiliary information (e.g., Wand and Jones (1995)). Another self-suggesting approach to using auxiliary data to produce model-assisted kernel estimators for $f(y)$ is to directly use regression models to model the relationship between the auxiliary variable X and the whole kernel $K(h^{-1}\{y-Y_i\})$ instead of modeling the relationship between X and Y. In the current chapter, we use this later approach to propose a third model-assisted kernel density estimator for $f(y)$. Interestingly, it is found that this new estimator preserves the same bandwidth order as in classical kernel density estimators $O(n^{-1/5})$ (see Section 4.5). The modeling step under the new approach that we discussed above is detailed in Section 4.2. The third proposed model-assisted density estimator is introduced in the same Section. The statistical properties of the new estimator are investigated in Sections 4.3 and 4.4. In Section 4.5, we consider the bandwidth selection problem for the new density estimator. Proofs of some technical lemmas are collected in Section 4.6.

4.2. Model-Assisted KDE Using Nonparametric Regression Models of A Kernel-Transformed Study Variable

Let $Z_i = K(h^{-1}\{y-Y_i\})$, for $i \in U$, define a kernel-transformed study variable. Note that the form of the transformation can be chosen freely because we have full control over the kernel function. Most likely we do not know a specific parametric form for the relationship between the transformed study variable Z and the auxiliary variable X. Thus, we resort to the nonparametric regression technique, used in Section 3.2, to model such relationship:

$$Z_i = m(X_i;y) + \sigma(X_i;y)\varepsilon_i, \quad i = 1,2,\ldots,N, \tag{4.2}$$

where $m(x;y) = E_\xi(Z|X=x)$ and $\sigma^2(x;y) = V_\xi(Z|X=x)$ are unspecified smooth functions of x and the variables ε_i are IID with zero mean and unit variance. Note that the point y, at which the density f is to be estimated, is part of the definition of the new study variable Z and, thus, it appears in both regression and variance functions in model (4.2). These functions also depend on the bandwidth, h, to be used for the new density estimator (see Eq. (4.7)) since h is part of Z. This dependence can be made explicit by writing $m(x;y,h)$

and $\sigma(x;y,h)$. In what follows, we tend to use the shorter notations $m(x;y)$ and $\sigma(x;y)$ unless we want to emphasize the dependence of these functions on h. Based on the entire finite population, the p-th degree local polynomial estimator of $m(\cdot;y)$ at the point x_i is (see Section 3.2)

$$m_U(x_i;y,p,b) = \mathbf{e}_1^T \left[\mathbf{X}_{iU}^T \mathbf{W}_{iU} \mathbf{X}_{iU}\right]^{-1} \mathbf{X}_{iU}^T \mathbf{W}_{iU} \mathbf{z}_U, \tag{4.3}$$

where $\mathbf{X}_{iU}$ is an $N \times (p+1)$ design matrix defined in Section 3.2, $\mathbf{W}_{iU}$ is an $N \times N$ kernel-weighting matrix having the form $\mathbf{W}_{iU} = diag\left\{\psi_b(X_j - x_i)\right\}_{j \in U}$ with ψ being a kernel function and b a smoothing parameter, $\mathbf{z}_U = (Z_1\ Z_2\ \ldots\ Z_N)^T$ and the vector $\mathbf{e}_1$ is as defined in Section 3.2. The estimator $m_U(\cdot;y,p,b)$ is not obtainable as it requires observing Y for the entire finite population. Therefore, we resort to the following sample-based local polynomial estimator of $m(x_i;y)$:

$$\hat{m}(x_i;y,p,b) = \mathbf{e}_1^T \left[\mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{X}_{is}\right]^{-1} \mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{z}_s, \tag{4.4}$$

where $\mathbf{X}_{is}$ is as defined in Section 3.2, $\mathbf{W}_{is}$ is an $n \times n$ sample weighting matrix of the form $\mathbf{W}_{is} = diag\left\{d_j \psi_b(X_j - x_i)\right\}_{j \in s}$ and $\mathbf{z}_s$ is the sample analog of $\mathbf{z}_U$. An important special case of $m_U(x;y,p,b)$ and $\hat{m}(x;y,p,b)$ is the local linear estimator obtained by setting $p = 1$:

$$m_U(x;y,1,b) = \sum_{j \in U} \frac{\{T_2(x;b) - (X_j - x)T_1(x;b)\}\psi_b(X_j - x)Z_j}{T_2(x;b)T_0(x;b) - T_1^2(x;b)} = \sum_{j \in U} l_j(x)Z_j, \tag{4.5}$$

and

$$\hat{m}(x;y,1,b) = \sum_{j \in s} \frac{d_j\{\hat{T}_2(x;b) - (x_j - x)\hat{T}_1(x;b)\}\psi_b(x_j - x)z_j}{\hat{T}_2(x;b)\hat{T}_0(x;b) - \hat{T}_1^2(x;b)} = \sum_{j \in s} l_{js}(x)z_j, \tag{4.6}$$

where $T_r(x;b) = \sum_{k \in U}(X_k - x)^r \psi_b(X_k - x)$ and $\hat{T}_r(x;b) = \sum_{k \in s} d_k (x_k - x)^r \psi_b(x_k - x)$ for $r = 0, 1, 2$. Following the recommendations of Wand and Jones (1995) and Fan and Gijbels (1996) regarding the choice of the polynomial degree (see Section 3.2), we will restrict our attention to local polynomial fits of degree $p = 1$, i.e., local linear fits.

4.2.1. Proposed Estimator

Using the modeling described in the previous section, we propose the following model-assisted kernel density estimator for $f(y)$:

$$\hat{f}_d(y;h) = \frac{1}{Nh}\left[\sum_{i\in s} d_i \left\{K\left(\frac{y-y_i}{h}\right) - \hat{m}(x_i;y,1,b)\right\} + \sum_{i\in U}\hat{m}(x_i;y,1,b)\right], \tag{4.7}$$

where $\hat{m}(x_i;y,1,b)$ is given by (4.6). The estimator $\hat{f}_d(y)$ enjoys the same two properties we mentioned for the estimator $\hat{f}_{dn}(y)$ right after Eq. (3.5), namely; (1) it does not assume any parametric form for the relationship between X and Y; (2) it exploits the design information, through using the design weights, causing the estimator to have desirable design-based properties as will be shown latter in this Section.

4.2.2. Main Assumptions

In addition to Assumptions A.1, A.2(i–ii) and A.4 in Section 2.2.2 and Assumption B.1 in Section 3.2.2, the following set of assumptions are required for the derivation of our results.

C.1 *(The bandwidths)*:

(i) $h_\tau(n_\tau, N_\tau) = h_\tau = h$ is such that $h_\tau \to 0$, $n_\tau h_\tau \to \infty$ and $n_\tau h_\tau^2 \to \infty$ as $\tau \to \infty$.

(ii) $b_\tau(n_\tau, N_\tau) = b_\tau = b$ is such that $b_\tau \to 0$, $n_\tau b_\tau \to \infty$ and $h_\tau = O(b_\tau)$ as $\tau \to \infty$.

C.2 *(Regression and variance functions)*:

(i) The regression function $m(x;y)$ has a bounded continuous second derivative.

(ii) The variance function $\sigma^2(x;y)$ is bounded and continuous.

C.3 *(Inclusion probabilities)*: The inclusion probabilities produced by the sampling design $\mathcal{P}(\cdot)$ are assumed to satisfy the following:

(i) $\lim_{\tau\to\infty} n_\tau^2 \max_{(i_1,i_2,i_3,i_4)\in S_{4,\tau}} |E_{\mathcal{P}}[(I_{i_1}-\pi_{i_1})(I_{i_2}-\pi_{i_2})(I_{i_3}-\pi_{i_3})(I_{i_4}-\pi_{i_4})]| < \infty$.

(ii) $\limsup_{\tau\to\infty} n_\tau \max_{(i_1,i_2,i_3)\in S_{3,\tau}} \left| E_{\mathcal{P}}\left[(I_{i_1}-\pi_{i_1})^2(I_{i_2}-\pi_{i_2})(I_{i_3}-\pi_{i_3})\right]\right| < \infty.$

(iii) $\lim_{\tau\to\infty} \max_{(i_1,i_2,i_3,i_4)\in S_{4,\tau}} \left| E_{\mathcal{P}}\left[(I_{i_1}I_{i_2}-\pi_{i_1 i_2})(I_{i_3}I_{i_4}-\pi_{i_3 i_4})\right]\right| = 0,$

where $S_{l,\tau}$ denotes the set of all distinct l-tuples $(i_1,\ldots,i_l)$ from U_τ. These design assumptions hold for many sampling designs such as simple random sampling without replacement and stratified sampling with fixed stratum boundaries (e.g., Breidt and Opsomer (2000)).

4.3. Properties of $\hat{f}_d(y)$

First, we introduce the following three general properties of the estimator $\hat{f}_d(y)$.

Property 1. The proposed estimator, $\hat{f}_d(y)$, is a weighted kernel density estimator where the weights are the inverse inclusion probabilities suitably modified to use the auxiliary information;

$$\hat{f}_d(y) = \frac{1}{Nh}\sum_{i\in s} w_{is} K\left(\frac{y-y_i}{h}\right), \tag{4.8}$$

where $w_{is} = d_i + \sum_{j\in U}(1-I_j d_j) l_{is}(x_j)$.

Proof: From (4.7), we can write

$$\begin{aligned}
\hat{f}_d(y) &= \frac{1}{Nh}\left[\sum_{i\in s} d_i K\left(\frac{y-y_i}{h}\right) - \sum_{j\in s} d_j \hat{m}(x_j;y,1,b) + \sum_{j\in U}\hat{m}(x_j;y,1,b)\right]\\
&= \frac{1}{Nh}\left[\sum_{i\in s} d_i K\left(\frac{y-y_i}{h}\right) + \sum_{j\in U}(1-I_j d_j)\hat{m}(x_j;y,1,b)\right]\\
&= \frac{1}{Nh}\left[\sum_{i\in s} d_i K\left(\frac{y-y_i}{h}\right) + \sum_{j\in U}(1-I_j d_j)\sum_{i\in s} l_{is}(x_j) Z_i\right]\\
&= \frac{1}{Nh}\sum_{i\in s}\left[d_i + \sum_{j\in U}(1-I_j d_j) l_{is}(x_j)\right] K\left(\frac{y-y_i}{h}\right).
\end{aligned}$$

It can be shown that

$$\frac{1}{N}\sum_{i\in s} w_{is} = 1 \quad \text{and} \quad \frac{1}{N}\sum_{i\in s} w_{is} x_i = \frac{1}{N}\sum_{i\in U} X_i,$$

and, hence, the weights w_{is} are calibrating weights of first order. $\square$

Property 2. The proposed estimator in (4.7) is a genuine probability density function with probability approaching 1.

Proof: First, integrating $\hat{f}_d(y)$ over the real line gives

$$\int_{\mathbb{R}} \hat{f}_d(y)dy = \frac{1}{Nh}\sum_{i\in s} w_{is} \int_{\mathbb{R}} K\left(\frac{y-y_i}{h}\right) dy = \frac{1}{Nh}\sum_{i\in s} w_{is} h \int_{\mathbb{R}} K(u)du = 1,$$

by the change of variables $u = (y-y_i)/h$ and since $\sum_{i\in s} w_{is} = N$. Second, $\hat{f}_d(y) \geq 0$ for all $y \in \mathbb{R}$ with probability approaching 1 since $\hat{f}_d(y)$ is a consistent estimator (in the MSE sense) for $f(y)$ (see Theorem 4.2). $\square$

Property 3. The mean of the proposed estimator in (4.7) is the local linear (LL) estimator of the finite population mean $\bar{Y}_U$:

$$\hat{\bar{Y}}_{U,LL} = \int_{\mathbb{R}} y\hat{f}_d(y)dy = \frac{1}{N}\left[\sum_{i\in s} d_i\{y_i - \hat{r}(x_i;1,b)\} + \sum_{i\in U} \hat{r}(x_i;1,b)\right], \tag{4.9}$$

where $\hat{r}(x_i;1,b)$ has the same form as $\hat{m}(x_i;y,1,b)$, given in (4.6), after replacing z_j by y_j for all $j \in s$. That is, $\hat{r}(x_i;1,b)$ is the local linear fit of $r(x_i)$ in the model $Y_i = r(X_i) + u_i$.

Proof: From (4.8), we have

$$\begin{aligned}
\int_{\mathbb{R}} y\hat{f}_d(y)dy &= \frac{1}{Nh}\sum_{i\in s} w_{is} \int_{\mathbb{R}} yK\left(\frac{y-y_i}{h}\right) dy \\
&= \frac{1}{Nh}\sum_{i\in s} w_{is} \int_{\mathbb{R}} (y_i + hu)K(u)hdu \\
&= \frac{1}{N}\sum_{i\in s} w_{is} y_i \\
&= \frac{1}{N}\sum_{i\in s}\{d_i + \sum_{j\in U}(1 - I_j d_j)l_{is}(x_j)\}y_i \\
&= \frac{1}{N}\sum_{i\in s}\left[d_i y_i - \sum_{j\in s} d_j l_{is}(x_j)y_i + \sum_{j\in U} l_{is}(x_j)y_i\right] \\
&= \frac{1}{N}\left[\sum_{i\in s} d_i y_i - \sum_{j\in s} d_j \sum_{i\in s} l_{is}(x_j)y_i + \sum_{j\in U}\sum_{i\in s} l_{is}(x_j)y_i\right],
\end{aligned}$$

where we used the change of variables $u=(y-y_i)/h$ to get the second equality and the third equality follows from Assumptions A.3(i–ii). $\square$

Property 4. Integrating the estimator $\hat{f}_d(y)$ over the interval $(-\infty,t]$ produces the following model-assisted estimator of the distribution function $F(y)$:

$$\hat{F}_{LLS}(t)=\frac{1}{N}\left\{\sum_{j\in U}\hat{\mathcal{M}}(X_j;t)+\sum_{i\in s}d_i\mathcal{K}\left(\frac{t-y_i}{h}\right)-\sum_{j\in s}d_j\hat{\mathcal{M}}(x_j;t)\right\}, \tag{4.10}$$

where $\hat{\mathcal{M}}(x;t)=\sum_{i\in s}l_{is}(x)\mathcal{Z}_i$, with $l_{is}(x)$ being the local linear weights in (4.6), $\mathcal{Z}_i=\mathcal{K}(h^{-1}\{t-y_i\})$ and $\mathcal{K}(y)=\int_{-\infty}^{y}K(u)du$, is the lcoal linear fit of $\mathcal{M}(x;t)=E_\xi[\mathcal{Z}|X=x]$. The estimator $\hat{F}_{LLS}(y)$ is a smooth version of the model-assisted estimator suggested by Johnson et al. (2008) to estimate the finite population CDF $F_U(y)$ (see Section 1.5).

Proof: From (4.8), we can write

$$\begin{aligned}
&\int_{-\infty}^{t}\hat{f}_d(y)dy\\
&=\frac{1}{Nh}\sum_{i\in s}w_{is}\int_{-\infty}^{t}K\left(\frac{y-y_i}{h}\right)dy\\
&=\frac{1}{N}\sum_{i\in s}w_{is}\int_{-\infty}^{\frac{t-y_i}{h}}K(u)du\\
&=\frac{1}{N}\sum_{i\in s}w_{is}\mathcal{K}\left(\frac{t-y_i}{h}\right)\\
&=\frac{1}{N}\sum_{i\in s}\{d_i+\sum_{j\in U}(1-I_jd_j)l_{is}(x_j)\}\mathcal{K}\left(\frac{t-y_i}{h}\right)\\
&=\frac{1}{N}\left[\sum_{j\in U}\sum_{i\in s}l_{is}(x_j)\mathcal{K}\left(\frac{t-y_i}{h}\right)+\sum_{i\in s}d_i\mathcal{K}\left(\frac{t-y_i}{h}\right)-\sum_{j\in s}d_j\sum_{i\in s}l_{is}(x_j)\mathcal{K}\left(\frac{t-y_i}{h}\right)\right].
\end{aligned}$$

Noting that $\hat{\mathcal{M}}(x;t)=\sum_{i\in s}l_{is}(x)\mathcal{K}\left(\frac{t-y_i}{h}\right)$ completes the proof. $\square$

The design-based properties of $\hat{f}_d(y;h)$ are summarized in the following theorem.

Theorem 4.1. *Given the set of Assumptions in Section 4.2.2, the estimator $\hat{f}_d(y;h)$ is asymptotically design-unbiased in the sense that*

$$\lim_{\tau\to\infty}E_{\mathcal{P}}\left[\hat{f}_d(y;h)-f_U(y;h)\right]=0 \qquad \textit{with } \xi\textit{-probability 1},$$

and is design-consistent in the sense that for all $\gamma > 0$,

$$\lim_{\tau\to\infty} E_{\mathcal{P}}\left[I(|\hat{f}_d(y;h) - f_U(y;h)| > \gamma)\right] = 0 \qquad \textit{with } \xi\textit{-probability 1},$$

where $I(A)$ equals one if event A holds and zero otherwise.

Proof: The proof is along the lines of Theorem 1 of Robinson and Särndal (1983). First note that, by the Markov inequality, the theorem follows if we show that

$$\lim_{\tau\to\infty} E_{\mathcal{P}}|\hat{f}_d(y) - f_U(y)| = 0. \tag{4.11}$$

Towards this, we write

$$\begin{aligned}\hat{f}_d(y) - f_U(y) = \frac{1}{Nh}\sum_{i\in U}\left\{K\left(\frac{y-y_i}{h}\right) - m_U(x_i;y)\right\}(I_i d_i - 1)\\ +\frac{1}{Nh}\sum_{i\in U}\left\{\hat{m}(x_i;y) - m_U(x_i;y)\right\}(1 - I_i d_i).\end{aligned}$$

Therefore,

$$\begin{aligned}E_{\mathcal{P}}|\hat{f}_d(y) - f_U(y)| &\le E_{\mathcal{P}}\left|\frac{1}{Nh}\sum_{i\in U}\left\{K\left(\frac{y-y_i}{h}\right) - m_U(x_i;y)\right\}(I_i d_i - 1)\right|\\ &\quad +E_{\mathcal{P}}\left|\frac{1}{Nh}\sum_{i\in U}\left\{\hat{m}(x_i;y) - m_U(x_i;y)\right\}(1 - I_i d_i)\right|\\ &:= \kappa_1 + \kappa_2.\end{aligned} \tag{4.12}$$

Notice that

$$\begin{aligned}\kappa_1 &\le E_{\mathcal{P}}\left[\frac{1}{Nh}\sum_{i\in U}\left\{K\left(\frac{y-y_i}{h}\right) - m_U(x_i;y)\right\}(I_i d_i - 1)\right]^2\\ &= E_{\mathcal{P}}\left[\frac{1}{N^2h^2}\sum_{i\in U}\left\{Z_i - m_U(x_i;y)\right\}^2 (I_i d_i - 1)^2\right]\\ &\quad +E_{\mathcal{P}}\left[\frac{1}{N^2h^2}\sum_{i,j\in U, i\neq j}\left\{Z_i - m_U(x_i;y)\right\}\left\{Z_j - m_U(x_j;y)\right\}(I_i d_i - 1)(I_j d_j - 1)\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N^2h^2}\sum_{i\in U}\{Z_i - m_U(x_i;y)\}^2 E_{\mathcal{P}}(I_i-\pi_i)^2/\pi_i^2 \\
&\quad + \frac{1}{N^2h^2}\sum_{i,j\in U, i\neq j}\{Z_i - m_U(x_i;y)\}\{Z_j - m_U(x_j;y)\} E_{\mathcal{P}}\{(I_i-\pi_i)(I_j-\pi_j)\}/\pi_i\pi_j \\
&= \frac{1}{N^2h^2}\sum_{i\in U}\{Z_i - m_U(x_i;y)\}^2 (1-\pi_i)/\pi_i \\
&\quad + \frac{1}{N^2h^2}\sum_{i,j\in U, i\neq j}\{Z_i - m_U(x_i;y)\}\{Z_j - m_U(x_j;y)\}(\pi_{ij}-\pi_i\pi_j)/\pi_i\pi_j \\
&\leq \frac{1}{Nh^2\lambda}\left[\frac{1}{N}\sum_{i\in U}\{Z_i - m_U(x_i;y)\}^2\right] \\
&\quad + \frac{1}{N^2h^2}\sum_{i,j\in U, i\neq j}\{Z_i - m_U(x_i;y)\}\{Z_j - m_U(x_j;y)\}(\pi_{ij}-\pi_i\pi_j)/\pi_i\pi_j \\
&:= \kappa_{11} + \kappa_{12}.
\end{aligned}
$$

Now, $\kappa_{11} \to 0$ as $\tau \to \infty$ because Z_i is bounded for all i by the boundedness of the kernel, $m_U(x_i;y)$ is bounded in i assuming the denominator of $m_U(x_i;y)$ to be bounded away from zero, and the assumptions that $n_\tau h_\tau^2 \to \infty$ and $n_\tau/N_\tau \to \pi$ as $\tau \to \infty$. Further,

$$
\begin{aligned}
\kappa_{12} &\leq \frac{1}{nN^2h^2}\sum_{i,j\in U, i\neq j}\{Z_i - m_U(x_i;y)\}\{Z_j - m_U(x_j;y)\}\, n \max_{i,j\in U,i\neq j}|\pi_{ij}-\pi_i\pi_j|\lambda^{-2} \\
&= \frac{\lambda^{-2}}{nh^2}\left[n \max_{i,j\in U,i\neq j}|\pi_{ij}-\pi_i\pi_j|\right]\left[\frac{1}{N^2}\sum_{i,j\in U, i\neq j}\{Z_i - m_U(x_i;y)\}\{Z_j - m_U(x_j;y)\}\right] \\
&\leq \frac{\lambda^{-2}}{nh^2}\left[n \max_{i,j\in U,i\neq j}|\pi_{ij}-\pi_i\pi_j|\right]\left[\frac{1}{N}\sum_{i\in U}\{Z_i - m_U(x_i;y)\}^2\right] \to 0, \quad \text{as} \quad \tau\to\infty,
\end{aligned}
$$

by assumptions A.4 and C.1(i). Thus, $\kappa_1 \to 0$ as $\tau \to \infty$. Next, we consider κ_2. Observe that

$$
\begin{aligned}
\kappa_2 &\leq \frac{1}{Nh}\sum_{i\in U} E_{\mathcal{P}}\left|\{\hat{m}(x_i;y) - m_U(x_i;y)\}(1-I_id_i)\right| \\
&\leq \frac{1}{Nh}\sum_{i\in U}\left[E_{\mathcal{P}}\{\hat{m}(x_i;y) - m_U(x_i;y)\}^2 E_{\mathcal{P}}\{(1-I_id_i)\}^2\right]^{1/2} \\
&= \frac{1}{Nh}\sum_{i\in U}\left[E_{\mathcal{P}}\{\hat{m}(x_i;y) - m_U(x_i;y)\}^2 (1-\pi_i)/\pi_i\right]^{1/2} \\
&= \frac{\lambda^{-1/2}}{Nh}\sum_{i\in U}\left[E_{\mathcal{P}}\{\hat{m}(x_i;y) - m_U(x_i;y)\}^2\right]^{1/2} \to 0 \quad \text{as} \quad \tau\to\infty,
\end{aligned}
$$

as $\hat{m}(x;y)$ is design-consistent for $m_U(x;y)$. Substituting these results into (4.12) completes the proof. $\square$

We need the following two lemmas to prove Theorem 4.2 which gives both the bias and the MISE formulae of the estimator $\hat{f}_d(\cdot)$ under the combined approach for inference.

Lemma 4.1. *Under Assumptions C.1(ii) and C.2,*

$$E_\xi\left[\{m_U(X;y)-m(X;y)\}^2|X_1,\ldots,X_N\right]$$
$$=\frac{1}{4}c_\psi^2\left\{\frac{\partial^2}{\partial x^2}m(x;y)\right\}^2 b^4+\frac{d_\psi}{Nb}\frac{\sigma^2(x;y)}{g(x)}+o_p\left(b^4+\frac{1}{Nb}\right). \tag{4.13}$$

Proof: See Section 4.6.

Lemma 4.2. *Under the set of Assumptions in Section 4.2.2,*

$$\frac{1}{N^2h^2}E_C\left[\sum\sum_{i,j\in U}(\hat{m}(X_i;y)-m_U(X_i;y))(\hat{m}(X_j;y)-m_U(X_j;y))(I_id_i-1)(I_jd_j-1)\right]$$
$$=o_p\left(\frac{1}{nh}\right). \tag{4.14}$$

Proof: See Section 4.6.

Theorem 4.2. *Suppose the set of Assumptions in Section 4.2.2 hold. Then, under the combined mode of inference, the bias and the MISE of the estimator $\hat{f}_d(\cdot;h)$ are;*

$$Bias_C\left[\hat{f}_d(y;h)\right]=\frac{1}{2}h^2c_Kf''(y)+o\left(h^2\right), \tag{4.15}$$

and

$$MISE_C\left[\hat{f}_d(\cdot;h)\right]=(nh)^{-1}\delta^*d_K+\frac{1}{4}h^4c_K^2d_{f''}+o\left(h^4+\frac{1}{Nh}\right), \tag{4.16}$$

where $\delta^*=\left[nN^{-2}\sum_{i\in U}\Delta_i+n/N\right]$.

Remarks: The following are some conclusions to be drawn from Theorem 4.2.

(i) The leading term in the bias of the estimator $\hat{f}_d(y)$ is identical to the leading term of the bias of the classical model-based kernel density estimator that does not use any auxiliary information. This is the same conclusion we had for the two model-assisted estimators we proposed in Chapters 2 and 3.

(ii) The leading term in the variance of the estimator $\hat{f}_d(y)$ is similar to the leading term of the variance of classical model-based kernel density estimators but the former has a component that adjusts for the sampling design effect; $\delta^* = \left[nN^{-2}\sum_{i\in U}\Delta_i + nN^{-1}\right]$.

(iii) Although, the effect of the auxiliary data, used in the estimator $\hat{f}_d(y)$, does not appear explicitly in the leading terms of the MISE, it is contained in the small-order terms $o[h^4+(1/Nh)]$. This fact is realized from the details of the proof of Theorem 4.2 where it becomes clear that the quantities $f(y|x)$, $(\partial^2/\partial x^2)f(y|x)$ and $(\partial^2/\partial y^2)f(y|x)$ are present in the terms collected in the order statement $o[h^4+(1/Nh)]$. More precisely, we have

$$
\begin{aligned}
&o\left(h^4+\frac{1}{Nh}\right)\\
&=h^4\left[\frac{c_\psi^2}{4N}\int_{\mathbb{R}}\left(\frac{\partial^2}{\partial x^2}f(y|x)\right)^2 g(x)dx+\frac{1}{n}\left\{\int_{\mathbb{R}}\left(\frac{\partial^2}{\partial x^2}f(y|x)\right)g(x)dx\right\}^2\right]+\frac{1}{Nh}\times\\
&\quad\left[h\left\{c_{K^2}\int_{\mathbb{R}}\left(\frac{\partial^2}{\partial y^2}f(y|x)\right)g(x)dx-\int_{\mathbb{R}}f^2(y|x)dx\right\}+\frac{1}{Nh}d_\psi d_K\int_{\mathbb{R}}f(y|x)dx\right].
\end{aligned} \tag{4.17}
$$

Proof: First note that,

$$
\begin{aligned}
&\hat{f}_d(y;h)-f(y)\\
&=\frac{1}{Nh}\sum_{i\in U}\left\{K\left(\frac{y-y_i}{h}\right)-m(x_i;y)\right\}(I_id_i-1)+\frac{1}{N}\sum_{i\in U}\left\{\frac{1}{h}K\left(\frac{y-y_i}{h}\right)-f(y)\right\}\\
&\quad+\frac{1}{Nh}\sum_{i\in U}\{m_U(x_i;y)-\hat{m}(x_i;y)\}(I_id_i-1)+\frac{1}{Nh}\sum_{i\in U}\{m(x_i;y)-m_U(x_i;y)\}(I_id_i-1)\\
&:=J_1+J_2+J_3+J_4.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
MSE_C\left[\hat{f}_d(y;h)\right] &= E_C\left[\hat{f}_d(y;h) - f(y)\right]^2 \\
&= E_C\left[J_1 + J_2 + J_3 + J_4\right]^2 \\
&= \sum_{k=1}^{4} E_C(J_k^2) + \sum\sum_{k \neq l} E_C(J_k J_l). \qquad (4.18)
\end{aligned}
$$

We separately work each expectation in (4.18) as follows:

$$
\begin{aligned}
E_C(J_1^2) &= \frac{1}{N^2h^2} E_C\left[\sum_{i\in U}\left(Z_i - m(x_i;y)\right)\left(I_i d_i - 1\right)\right]^2 \\
&= \frac{1}{N^2h^2} E_\xi\left\{E_{\mathcal{P}}\left[\sum_{i\in s} d_i\left(Z_i - m(x_i;y)\right) - \sum_{i\in U}\left(Z_i - m(x_i;y)\right)\right]^2\right\} \\
&= \frac{1}{N^2h^2} E_\xi\left[\sum\sum_{i,j\in U}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\left(Z_i - m(X_i;y)\right)\left(Z_j - m(X_j;y)\right)\right] \\
&= \frac{1}{N^2h^2} E_\xi\left[\sum_{i\in U}\Delta_i\left(Z_i - m(X_i;y)\right)^2 + \sum\sum_{i\neq j\in U}\Delta_{ij}\left(Z_i - m(X_i;y)\right)\left(Z_j - m(X_j;y)\right)\right]. \qquad (4.19)
\end{aligned}
$$

But

$$
\begin{aligned}
&\frac{1}{Nh^2} E_\xi\left[\left(Z_i - m(X_i;y)\right)^2 |X_i\right] \\
&\quad= \frac{1}{Nh^2} E_\xi\left[\left(K\left(\frac{y-Y_i}{h}\right) - E_\xi\left\{K\left(\frac{y-Y_i}{h}\right)|X_i\right\}\right)^2 |X_i\right] \\
&\quad= \frac{1}{Nh^2} E_\xi\left[\left(K\left(\frac{y-Y_i}{h}\right) - E_\xi\left\{K\left(\frac{y-Y_i}{h}\right)|X_i\right\}\right)^2 |X_i\right] \\
&\quad= \frac{1}{Nh^2}\left[E_\xi\left(K^2\left(\frac{y-Y_i}{h}\right)|X_i\right) - \left(E_\xi\left\{K\left(\frac{y-Y_i}{h}\right)|X_i\right\}\right)^2\right],
\end{aligned}
$$

$$
\begin{aligned}
E_\xi\left[K^2\left(\frac{y-Y_i}{h}\right)|X_i\right] &= \int_{\mathbb{R}} K^2\left(\frac{y-y_i}{h}\right) f(y_i|x_i)\,dy_i \\
&= h\int_{\mathbb{R}} K^2(u) f(y-hu|x_i)\,du
\end{aligned}
$$

$$
\begin{aligned}
&= h \int_{\mathbb{R}} K^2(u) \left[f(y|x_i) - huf'(y|x_i) + h^2u^2 f''(y|x_i) + \dots \right] du \\
&= hd_K f(y|x_i) - h^2 \mu(K^2) f'(y|x_i) + o_p(h^2) \\
&= hd_K f(y|x_i) + o_p(h^2),
\end{aligned}
$$

since $\mu(K^2) = \int_{\mathbb{R}} uK^2(u)du = 0$ by assumption A.2 and we used the notation $f^{(s)}(y|x) \equiv (\partial^s/\partial y^s) f(y|x)$, and

$$
\begin{aligned}
E_\xi \left[K\left(\frac{y-Y_i}{h}\right) |X_i \right] &= \int_{\mathbb{R}} K\left(\frac{y-y_i}{h}\right) f(y_i|x_i)dy_i \\
&= h \int_{\mathbb{R}} K(u) f(y-hu|x_i)du \\
&= h \int_{\mathbb{R}} K(u) \left[f(y|x_i) - huf'(y|x_i) + h^2u^2 f''(y|x_i) + \dots \right] du \\
&= hf(y|x_i) + o_p(h^2).
\end{aligned}
$$

Thus,

$$
\frac{1}{Nh^2} E_\xi \left[(Z_i - m(X_i;y))^2 |X_i \right] = \frac{1}{Nh} d_K f(y|x_i) + o_p\left(\frac{1}{Nh}\right). \tag{4.20}
$$

Additionally, by the independence of the pairs (X_i, Y_i) in the finite population

$$
\begin{aligned}
&\frac{1}{h^2} E_\xi \left[(Z_i - m(X_i;y)) \left(Z_j - m(X_j;y)\right) |X_i, X_j \right] \\
&\quad = \frac{1}{h^2} \left[E_\xi \left(\left\{ K\left(\frac{y-Y_i}{h}\right) - E_\xi \left(K\left(\frac{y-Y_i}{h}\right) |X_i \right) \right\} |X_i \right) \right. \\
&\qquad \left. \times E_\xi \left(\left\{ K\left(\frac{y-Y_j}{h}\right) - E_\xi \left(K\left(\frac{y-Y_j}{h}\right) |X_j \right) \right\} |X_i \right) \right] = 0.
\end{aligned} \tag{4.21}
$$

Taking expectations of the right-hand side in Eqs. (4.20) and (4.21) with respect to X and substituting the results into (4.19), we get

$$
\begin{aligned}
E_C(J_1^2) &= (Nh)^{-1} d_K \left(\frac{1}{N} \sum_{i \in U} \Delta_i \int_{\mathbb{R}} f(y|x_i) g(x_i) dx_i \right) + o\left(\frac{1}{Nh}\right) \\
&= (Nh)^{-1} \left(\frac{1}{N} \sum_{i \in U} \Delta_i \right) d_K f(y) + o\left(\frac{1}{Nh}\right).
\end{aligned} \tag{4.22}
$$

Note that in the first equality in (4.22), $\Delta_i = (1-\pi_i)/\pi_i$ is taken out of the integral because the inclusion probabilities are independent of X by assumption A.4.

Standard results on kernel density estimation (e.g., Wand and Jones (1995)) give

$$
\begin{aligned}
E_C(J_2^2) &= E_C\left[\frac{1}{Nh}\sum_{i\in U} K\left(\frac{y-Y_i}{h}\right) - f(y)\right]^2 \\
&= E_C\left[f_U(y;h) - f(y)\right]^2 \\
&= (Nh)^{-1} d_K f(y) + \frac{1}{4}h^4 c_K^2 \{f''(y)\}^2 + o\left(h^4 + \frac{1}{Nh}\right). \qquad (4.23)
\end{aligned}
$$

Considering J_4 and using some work similar to the work that led to (4.19), we have

$$
\begin{aligned}
E_C(J_4^2) &= \frac{1}{N^2h^2} E_C\left[\sum_{i\in U} \left(m(x_i;y) - m_U(x_i;y)\right)\left(I_i d_i - 1\right)\right]^2 \\
&= \frac{1}{N^2h^2} E_\xi \left\{\sum_{i\in U} \Delta_i \left(m_U(X_i;y) - m(X_i;y)\right)^2\right. \\
&\quad \left. + \sum\sum_{i\neq j\in U} \Delta_{ij} \left(m_U(X_i;y) - m(X_i;y)\right)\left(m_U(X_j;y) - m(X_j;y)\right)\right\}. \qquad (4.24)
\end{aligned}
$$

From Lemma 4.1, we have

$$
\begin{aligned}
&E_\xi\left[(m_U(x;y) - m(x;y))^2 | X_1, \ldots, X_N\right] \\
&\qquad = \frac{1}{4}c_\psi^2 \left\{\frac{\partial^2}{\partial x^2} m(x;y)\right\}^2 b^4 + \frac{d_\psi}{Nb}\frac{\sigma^2(x;y)}{g(x)} + o_p\left(b^4 + \frac{1}{Nb}\right). \qquad (4.25)
\end{aligned}
$$

Also, note that

$$
\begin{aligned}
m(x_i;y) &= E_\xi\left[K\left(\frac{Y_i - y}{h}\right) | X_i = x_i\right] \\
&= \int_{\mathbb{R}} K\left(\frac{y_i - y}{h}\right) f(y_i|x_i) dy_i \\
&= \int_{\mathbb{R}} K(u) f(y + hu|x_i) h du \\
&= h\int_{\mathbb{R}} K(u)\left[f(y|x_i) + huf'(y|x_i) + h^2u^2 f''(y|x_i) + \cdots\right] du \\
&= hf(y|x_i) + o_p(h^2) \qquad (4.26)
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma^2(x_i;y) &= V_\xi \left[K\left(\frac{Y_i - y}{h}\right) |X_i = x_i \right] \\
&= E_\xi \left[K^2\left(\frac{Y_i - y}{h}\right) |X_i = x_i \right] - \left[E_\xi \left(K\left(\frac{Y_i - y}{h}\right) |X_i = x_i \right) \right]^2 \\
&= \int_{\mathbb{R}} K^2\left(\frac{y_i - y}{h}\right) f(y_i|x_i) dy_i - \left[hf(y|x_i) + o_p(h^2) \right]^2 \\
&= \int_{\mathbb{R}} K^2(u) f(y + hu|x_i) h du - \left[hf(y|x_i) + o_p(h^2) \right]^2 \\
&= h \int_{\mathbb{R}} K^2(u) \left[f(y|x_i) + huf'(y|x_i) + \cdots \right] du - \left[hf(y|x_i) + o_p(h^2) \right]^2 \\
&= hd_K f(y|x_i) + h^2 f'(y|x_i) \int_{\mathbb{R}} uK^2(u) du - h^2 f^2(y|x_i) + o_p(h^2) \\
&= hd_K f(y|x_i) - h^2 f^2(y|x_i) + o_p(h^2). \qquad (4.27)
\end{aligned}
$$

Using (4.25), (4.26) and (4.27) together, we get

$$
\begin{aligned}
&\frac{1}{Nh^2} E_\xi \left[(m_U(X_i;y) - m(X_i;y))^2 | X_1, \ldots, X_N \right] \\
&\qquad = \frac{1}{4Nh^2} c_\psi^2 \left(\frac{\partial^2}{\partial x_i^2} f(y|x_i) \right)^2 h^2 b^4 \\
&\qquad\quad + \frac{d_\psi}{N^2 h^2 b} \left[hd_K f(y|x_i) - h^2 f^2(y|x_i) \right] / g(x) + o_p\left(\frac{1}{N}\right) \\
&\qquad = o_p\left(b^4 + \frac{1}{Nb} \right). \qquad (4.28)
\end{aligned}
$$

On the other hand, the first term on the right-hand side of the MSE in Lemma 4.1 (see Eq. (4.13)) is the squared conditional bias of the estimate $m_U(x;y)$. Thus,

$$
E_\xi \left[(m_U(X_i;y) - m(X_i;y)) | X_1, \ldots, X_N \right] = \frac{1}{2} c_\psi \left(\frac{\partial^2}{\partial x_i^2} m(x_i;y) \right) b^2 + o_p(b^2).
$$

Since the pairs (X_i, Y_i) are IID in the model space, we have

$$
\begin{aligned}
&\frac{1}{nh^2} E_\xi \left[(m_U(X_i;y) - m(X_i;y))(m_U(X_j;y) - m(X_j;y)) | X_1, \ldots, X_N \right] \\
&\qquad\qquad = \frac{1}{nh^2} \left[\frac{1}{4} c_\psi^2 \left(\frac{\partial^2}{\partial x_i^2} m(x_i;y) \right) \left(\frac{\partial^2}{\partial x_j^2} m(x_j;y) \right) b^4 + o_p(b^4) \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{nh^2}\left[\frac{1}{4}c_\psi^2\left(h\frac{\partial^2}{\partial x_i^2}f(y|x_i)\right)\left(h\frac{\partial^2}{\partial x_j^2}f(y|x_j)\right)b^4 + o_p(b^4)\right] \\
&= O_p(n^{-1}b^4), \qquad (4.29)
\end{aligned}
$$

where the second equality follows from (4.26). Now, use (4.28) and (4.29) in (4.24) to get

$$
E_C(J_4^2) = \frac{1}{N}\sum_{i\in U}\Delta_i\, o\left(b^4+\frac{1}{Nb}\right) + \frac{n}{N^2}\sum\sum_{i\neq j\in U}\Delta_{ij}\, O(n^{-1}b^4) = o\left(b^4+\frac{1}{Nb}\right), \qquad (4.30)
$$

where the second equality follows because

$$
\frac{1}{N}\sum_{i\in U}\Delta_i = N^{-1}\sum_{i\in U}(\frac{1}{\pi_i}-1) \leq N^{-1}\sum_{i\in U}(\frac{1}{\lambda}-1) = \frac{(1-\lambda)}{\lambda} = O(1) \qquad (4.31)
$$

and

$$
\frac{n}{N^2}\sum\sum_{i\neq j}\Delta_{ij} = nN^{-2}\sum\sum_{i\neq j}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} \leq N^{-2}\sum\sum_{i\neq j}\frac{n\, \max_{i\neq j}|\pi_{ij}-\pi_i\pi_j|}{\lambda^2} = O(1), \quad (4.32)
$$

where the inequalities in (4.31) and (4.32) follow from Assumption A.4.

Now, consider J_3^2 and note that

$$
\begin{aligned}
E_C(J_3^2) = \frac{1}{N^2h^2}E_C\Bigg[&\sum\sum_{i,j\in U}(\hat{m}(X_i;y)-m_U(X_i;y))(\hat{m}(X_j;y)-m_U(X_j;y))\times \\
&(I_id_i-1)(I_jd_j-1)|X_1,\ldots,X_N\Bigg] = o_p\left(\frac{1}{nh}\right), \qquad (4.33)
\end{aligned}
$$

where the last equality is the result of Lemma 4.2.

Next, we deal with expectations of cross products.

$$
\begin{aligned}
&E_C(J_1J_2) \\
&= \frac{1}{N^2h}E_C\left[\left(\sum_{i\in U}\left\{K(\frac{y-y_i}{h})-m(x_i;y)\right\}(I_id_i-1)\right)\left(\frac{1}{N}\sum_{i\in U}\left\{\frac{1}{h}K(\frac{y-y_i}{h})-f(y)\right\}\right)\right] \\
&= \frac{1}{N^2h}E_\xi\left[\left(\sum_{i\in U}\left\{K(\frac{y-y_i}{h})-m(x_i;y)\right\}E_{\mathcal{P}}(I_id_i-1)\right)\times\right.
\end{aligned}
$$

$$\left(\frac{1}{N}\sum_{i\in U}\left\{\frac{1}{h}K(\frac{y-y_i}{h})-f(y)\right\}\right)\Bigg] = 0, \tag{4.34}$$

since $E_{\mathcal{P}}(I_i d_i - 1) = \pi_i d_i - 1 = 0$. For all other cross product expectations, we apply the Cauchy-Schwarz inequality to show that they have smaller orders. From (4.22) and (4.33), we have

$$E_C(J_1J_3) \leq [E_C(J_1^2)E_C(J_3^2)]^{1/2} = o\left(\frac{1}{\sqrt{nNh^2}}\right). \tag{4.35}$$

From (4.22) and(4.30), we have

$$E_C(J_1J_4) \leq [E_C(J_1^2)E_C(J_4^2)]^{1/2} = o\left(\sqrt{\frac{b^4}{Nh}+\frac{1}{N^2bh}}\right) = o\left(\sqrt{\frac{h^3}{N}+\frac{1}{N^2h^2}}\right). \tag{4.36}$$

Similarly, $E_C(J_2J_3) \leq o\left(\sqrt{\frac{1}{nNh^2}+\frac{h^3}{n}}\right)$ and $E_C(J_2J_4) \leq o\left(\sqrt{h^8+\frac{1}{N^2h^2}+\frac{h^3}{n}}\right)$. Finally, (4.33) and (4.30) imply that $E_C(J_3J_4) \leq o\left(\sqrt{\frac{h^3}{n}+\frac{1}{nNh^2}}\right)$. These results on cross product expectations show that the second term on the right-hand side of (4.18) is of smaller order compared to the first term. Using this conclusion and the results in (4.22) and (4.23) to substitute in (4.18), and defining $\delta^* = \left[nN^{-2}\sum_{i\in U}\Delta_i + n/N\right]$, we get

$$MSE_C\left[\hat{f}_d(y;h)\right] = (nh)^{-1}\delta^* d_K f(y) + \frac{1}{4}h^4c_K^2\{f''(y)\}^2 + o\left(h^4+\frac{1}{Nh}\right).$$

Integrating this MSE over y gives the MISE of $\hat{f}_d(\cdot;h)$ and the proof is complete. □

4.4. Asymptotic Distribution of $\hat{f}_d(y)$

In this section, the asymptotic distribution of the estimator $\hat{f}_d(y;h)$ is derived under both the design-based inference framework and the combined mode of inference. The following lemma gives the asymptotic distribution of a standardized version of $\hat{f}_d(y;h)$ in the design space. The result of this lemma is useful for making inference about the finite population smooth $f_U(y;h)$ based on $\hat{f}_d(y;h)$.

Lemma 4.3. *Suppose the set of Assumptions in Section 4.2.2 hold. Further, suppose that $K(x) \leq M$ for all $x \in \mathbb{R}$ and $\sqrt{n}h \to \infty$ as $n \to \infty$. Then, under SRSWOR, we have*

$$\frac{\hat{f}_d(y;h) - f_U(y;h)}{\hat{\Lambda}_{\mathcal{P}}^{1/2}} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1), \tag{4.37}$$

where

$$\hat{\Lambda}_{\mathcal{P}} = \left(1 - \frac{n}{N}\right) \frac{\left[\sum_{i\in s}(u_i(h) - \hat{m}_i^*(h))^2 - n^{-1}\{\sum_{i\in s}(u_i(h) - \hat{m}_i^*(h))\}^2\right]}{n(n-1)}, \tag{4.38}$$

with $u_i(h) = K_h(y - y_i)$ and

$$\hat{m}_i^*(h) = \hat{m}_i^*(x_i; y, b) = \sum_{j\in s} \frac{d_j\{\hat{T}_2(x_i;b) - (x_j - x_i)\hat{T}_1(x_i;b)\}\psi_b(x_j - x_i)z_j^*}{\hat{T}_2(x_i;b)\hat{T}_0(x_i;b) - \hat{T}_1^2(x_i;b)}, \tag{4.39}$$

where $z_j^* = K_h(y - y_j)$ and $\hat{T}_r(x;b) = \sum_{k\in s} d_k(x_k - x)^r \psi_b(x_k - x)$ for $r = 0,1,2$.

Proof: First, consider the following pseudo estimator which is obtained via replacing the sample-based regression fits $\hat{m}_i = \hat{m}(x_i;y,b)$ by the finite population fits $m_{Ui} = m_U(x_i;y,b)$ in the definition of the original estimator $\hat{f}_d(y;h)$ (see Eqs. (4.5) and (4.6) for definitions of m_{Ui} and $\hat{m}_i$, respectively):

$$\begin{aligned}
\tilde{f}_d(y;h) &= \frac{1}{Nh}\left[\sum_{i\in s} d_i \left\{K\left(\frac{y-y_i}{h}\right) - m_U(x_i;y,b)\right\} + \sum_{i\in U} m_U(x_i;y,b)\right] \\
&= \frac{1}{N}\left[\sum_{i\in s} d_i\{u_i(h) - m_{Ui}^*(h)\} + \sum_{i\in U} m_{Ui}^*(h)\right],
\end{aligned}$$

where

$$m_{Ui}^*(h) = \sum_{j\in U} \frac{\{T_2(x_i;b) - (X_j - x_i)T_1(x_i;b)\}\psi_b(X_j - x_i)u_j(h)}{T_2(x_i;b)T_0(x_i;b) - T_1^2(x_i;b)} = \sum_{j\in U} l_j(x_i)u_j(h), \tag{4.40}$$

with $u_j(h) = K_h(y - y_j)$ and $T_r(x;b) = \sum_{k\in U}(X_k - x)^r \psi_b(X_k - x)$ for $r = 0,1,2$. Under SR-

SWOR, $d_i = N/n$ for all $i \in U$ and, hence,

$$\tilde{f}_d(y;h) = \frac{1}{n}\sum_{i\in s}\{u_i(h) - m^*_{Ui}(h)\} + \frac{1}{N}\sum_{i\in U} m^*_{Ui}(h) = \frac{1}{n}\sum_{i\in s} w^*_i(h), \tag{4.41}$$

where $w^*_i(h) = u_i(h) - m^*_{Ui}(h) + (1/N)\sum_{j\in U} m^*_{Uj}(h)$. The design-variance of $\tilde{f}_d(y;h)$, under the SRSWOR design, is

$$\Lambda_{\mathcal{P}} = \left(1 - \frac{n}{N}\right)\frac{\left[\sum_{i\in U}(u_i(h) - m^*_{Ui}(h))^2 - N^{-1}\{\sum_{i\in U}(u_i(h) - m^*_{Ui}(h))\}^2\right]}{n(N-1)}. \tag{4.42}$$

Since $\tilde{f}_d(y;h)$ is a sample mean, as shown in Eq. (4.41), to show that

$$\frac{\tilde{f}_d(y;h) - f_U(y;h)}{\Lambda_{\mathcal{P}}^{1/2}} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1), \tag{4.43}$$

we need to verify the following Lyapunov's condition (e.g., Thompson (1997, pg. 59)):

$$\left(1 - \frac{n}{N}\right)\sum_{i\in s} E_{\mathcal{P}}|w^*_i(h) - E_{\mathcal{P}}[w^*_i(h)]|^{2+\eta} = o\left(\left[n^2\left(\frac{N-1}{N}\right)\Lambda_{\mathcal{P}}\right]^{(2+\eta)/2}\right). \tag{4.44}$$

Note that,

$$\begin{aligned}
E_{\mathcal{P}}[w^*_i(h)] &= \frac{1}{N}\sum_{i\in U} w^*_i(h) \\
&= \frac{1}{N}\sum_{i\in U}[u_i(h) - m^*_{Ui}(h) + (1/N)\sum_{j\in U} m^*_{Uj}(h)] \\
&= \frac{1}{N}\sum_{i\in U} u_i(h) = \frac{1}{N}\sum_{i\in U} K_h(y - y_i) = f_U(y;h).
\end{aligned}$$

Therefore,

$$\begin{aligned}
|w^*_i(h) - E_{\mathcal{P}}[w^*_i(h)]| &= \left|u_i(h) - m^*_{Ui}(h) + (1/N)\sum_{j\in U} m^*_{Uj}(h) - \frac{1}{N}\sum_{i\in U} u_i(h)\right| \\
&= \left|[u_i(h) - m^*_{Ui}(h)] - \frac{1}{N}\sum_{j\in U}[u_j(h) - m^*_{Uj}(h)]\right| \\
&\leq \left|[\frac{M}{h} - 0] - \frac{1}{N}\sum_{i\in U}[0 - \frac{M}{h}]\right| = \frac{2M}{h},
\end{aligned}$$

since $m^*_{Ui}(h) = \sum_{j \in U} l_j(x_i) u_j(h) \leq (M/h) \sum_{j \in U} l_j(x_i) = M/h$.

$$
\begin{aligned}
\sum_{i \in s} E_{\mathcal{P}} \left| w_i^*(h) - E_{\mathcal{P}}[w_i^*(h)] \right|^3 &\leq \sum_{i \in s} E_{\mathcal{P}} \left| w_i^*(h) - E_{\mathcal{P}}[w_i^*(h)] \right|^2 \max_{i \in s} |w_i^*(h) - E_{\mathcal{P}}[w_i^*(h)]| \\
&= \frac{2M}{h} \sum_{i \in s} E_{\mathcal{P}} \left[w_i^*(h) - E_{\mathcal{P}}\{w_i^*(h)\} \right]^2 \\
&= \frac{2M}{h} \sum_{i \in s} E_{\mathcal{P}} \left[\{u_i(h) - m^*_{Ui}(h)\} - \frac{1}{N} \sum_{j \in U} \{u_j(h) - m^*_{Uj}(h)\} \right]^2 \\
&= \frac{2M}{h} \frac{n}{N} \sum_{i \in U} \left[\{u_i(h) - m^*_{Ui}(h)\} - \frac{1}{N} \sum_{j \in U} \{u_j(h) - m^*_{Uj}(h)\} \right]^2 \\
&= \frac{2M}{h} n^2 \left(\frac{N-1}{N} \right) \left(1 - \frac{n}{N} \right)^{-1} \Lambda_{\mathcal{P}}.
\end{aligned}
$$

Thus,

$$
\left(1 - \frac{n}{N} \right) \sum_{i \in s} E_{\mathcal{P}} \left| w_i^*(h) - E_{\mathcal{P}}[w_i^*(h)] \right|^3 \leq \frac{2M}{h} n^2 \left(\frac{N-1}{N} \right) \Lambda_{\mathcal{P}}, \tag{4.45}
$$

and

$$
\begin{aligned}
&\frac{\left(1 - \frac{n}{N}\right) \sum_{i \in s} E_{\mathcal{P}} \left| w_i^*(h) - E_{\mathcal{P}}[w_i^*(h)] \right|^3}{\left[n^2 \left(\frac{N-1}{N} \right) \Lambda_{\mathcal{P}} \right]^{3/2}} \\
&\leq \frac{2M}{h} n^2 \left(\frac{N-1}{N} \right) \Lambda_{\mathcal{P}} \left[n^2 \left(\frac{N-1}{N} \right) \Lambda_{\mathcal{P}} \right]^{-3/2} \\
&= \frac{2M}{nh} \left(\frac{N-1}{N} \right)^{-1/2} \Lambda_{\mathcal{P}}^{-1/2} \\
&= \frac{2M}{\sqrt{n}h} \left(1 - \frac{n}{N} \right)^{-1/2} \left[\frac{\sum_{i \in U} (u_i(h) - m^*_{Ui}(h))^2 - N^{-1} \{\sum_{i \in U} (u_i(h) - m^*_{Ui}(h))\}^2}{N} \right]^{-1/2} \\
&\to 0, \qquad \text{as } \sqrt{n}h \to \infty.
\end{aligned} \tag{4.46}
$$

Therefore, Lyapunov's condition (4.44) holds with $\eta = 1$ and the asymptotic result in (4.43) is proven. Next, observe that

$$
\begin{aligned}
\hat{f}_d(y;h) - f_U(y;h) &= \frac{1}{N} \sum_{i \in s} d_i \{u_i(h) - \hat{m}_i^*(h)\} + \frac{1}{N} \sum_{i \in U} \hat{m}_i^*(h) - \frac{1}{N} \sum_{i \in U} u_i(h) \\
&= \frac{1}{N} \sum_{i \in U} \{u_i(h) - m^*_{Ui}(h)\} (I_i d_i - 1)
\end{aligned}
$$

$$+\frac{1}{N}\sum_{i\in U}\{\hat{m}_i^*(h)-m_{Ui}^*(h)\}(1-I_id_i)$$
$$=[\tilde{f}_d(y;h)-f_U(y;h)]+\frac{1}{Nh}\sum_{i\in U}\{\hat{m}_i(h)-m_{Ui}(h)\}(1-I_id_i).$$

It can be shown that (see the proof of Lemma 4.2 in Section 4.6)

$$\frac{1}{N^2h^2}\sum\sum_{i,j\in U}E_{\mathcal{P}}\left[\{\hat{m}_i(h)-m_{Ui}(h)\}\left\{\hat{m}_j(h)-m_{Uj}(h)\right\}(1-I_id_i)(1-I_jd_j)\right]$$
$$=o_p\left(\frac{1}{nh}\right).$$

Consequently,

$$\hat{f}_d(y;h)-f_U(y;h)=[\tilde{f}_d(y;h)-f_U(y;h)]+o_p\left(\frac{1}{\sqrt{nh}}\right). \tag{4.47}$$

Using (4.43) and (4.47) in Slutsky's theorem, we get

$$\frac{\hat{f}_d(y;h)-f_U(y;h)}{\Lambda_{\mathcal{P}}^{1/2}}\xrightarrow{\mathscr{L}_{\mathcal{P}}}N(0,1). \tag{4.48}$$

To complete the proof of the lemma, it remains to show that $\hat{\Lambda}_{\mathcal{P}}$, given in (4.38), is a design-consistent estimator for $\Lambda_{\mathcal{P}}$, or equivalently,

$$|\hat{\Lambda}_{\mathcal{P}}-\Lambda_{\mathcal{P}}|\xrightarrow{P_{\mathcal{P}}}0, \text{ as } n \text{ increases.}$$

Define

$$\tilde{\gamma}_{\mathcal{P}}^2=\frac{1}{n-1}\sum_{i\in s}\left[(u_i^*(h)-m_{Ui}(h))-n^{-1}\{\sum_{i\in s}(u_i^*(h)-m_{Ui}(h))\}\right]^2 \tag{4.49}$$

and

$$\gamma_{\mathcal{P}}^2=\frac{1}{N}\sum_{i\in U}\left[(u_i^*(h)-m_{Ui}(h))-N^{-1}\{\sum_{i\in U}(u_i^*(h)-m_{Ui}(h))\}\right]^2, \tag{4.50}$$

where $u_i^*(h) = hu_i(h) = K(\{y - y_i\}/h)$. Note that, by the boundedness assumption on $K(\cdot)$,

$$0 < \gamma_{\mathcal{P}}^2 \leq \frac{1}{N} \sum_{i \in U} \left[(M - 0) - N^{-1} \{ \sum_{i \in U} (0 - M) \} \right]^2 = 4M^2 < \infty. \tag{4.51}$$

Therefore,

$$\max_{i \in U} \left[(u_i^*(h) - m_{Ui}(h)) - N^{-1} \{ \sum_{i \in U} (u_i^*(h) - m_{Ui}(h)) \} \right]^2 / N\gamma_{\mathcal{P}}^2 \leq \frac{4M^2}{N\gamma_{\mathcal{P}}^2} \to 0 \text{ as } N \to \infty. \tag{4.52}$$

The bounds in (4.51) and (4.52) imply that conditions (2.10) and (2.12) on page 294 of Sen (1988) are satisfied. Thus, using the result in (2.19) of Sen (1988), we have

$$|\tilde{\gamma}_{\mathcal{P}}^2 - \gamma_{\mathcal{P}}^2| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.} \tag{4.53}$$

Consequently, taking $\tilde{\Lambda}_{\mathcal{P}} = (1 - n/N)\tilde{\gamma}_{\mathcal{P}}^2/nh^2$, we have

$$|\tilde{\Lambda}_{\mathcal{P}} - \Lambda_{\mathcal{P}}| = \frac{1}{nh^2} \left(1 - \frac{n}{N}\right) \left| \tilde{\gamma}_{\mathcal{P}}^2 - \frac{N}{(N-1)} \gamma_{\mathcal{P}}^2 \right| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.} \tag{4.54}$$

Define

$$\hat{\gamma}_{\mathcal{P}}^2 = \frac{1}{n-1} \sum_{i \in s} \left[(u_i^*(h) - \hat{m}_i(h)) - n^{-1} \{ \sum_{i \in s} (u_i^*(h) - \hat{m}_i(h)) \} \right]^2,$$

and observe that the only difference between $\hat{\gamma}_{\mathcal{P}}^2$ and $\tilde{\gamma}_{\mathcal{P}}^2$ is that $\hat{\gamma}_{\mathcal{P}}^2$ depends on $\hat{m}(h)$ while $\tilde{\gamma}_{\mathcal{P}}^2$ depends on $m_U(h)$. Since, in the design space, $\hat{m}(h)$ consistently estimates $m_U(h)$, the results of Randles (1982) imply that $\hat{\gamma}_{\mathcal{P}}^2$ and $\tilde{\gamma}_{\mathcal{P}}^2$ share the same limiting properties. Additionally, it is readily seen that $\hat{\Lambda}_{\mathcal{P}} = (1 - n/N)\hat{\gamma}_{\mathcal{P}}^2/nh^2$. Therefore, (4.54) implies that

$$|\hat{\Lambda}_{\mathcal{P}} - \Lambda_{\mathcal{P}}| = \frac{1}{nh^2} \left(1 - \frac{n}{N}\right) \left| \hat{\gamma}_{\mathcal{P}}^2 - \frac{N}{(N-1)} \gamma_{\mathcal{P}}^2 \right| \xrightarrow{P_{\mathcal{P}}} 0, \text{ as } n \text{ increases.} \tag{4.55}$$

The proof of the lemma is complete upon using the results in (4.48) and (4.55) in Slutsky's theorem. $\square$

Theorem 4.3. *Suppose Assumptions A.1–A.4 and C.1–C.3 hold. Further, suppose that $K(x) \leq M$ for all $x \in \mathbb{R}$ and $\sqrt{n}h \to \infty$ as $n \to \infty$. Then, under SRSWOR, we have*

$$\frac{\hat{f}_d(y;h) - E_\xi\{f_U(y;h)\}}{\sqrt{V_C\{\hat{f}_d(y;h)\}}} \xrightarrow{\mathscr{L}_C} N(0,1), \tag{4.56}$$

where $E_\xi\{f_U(y;h)\}$ is given by (2.20) *and*

$$V_C\left[\hat{f}_d(y;h)\right] = (nh)^{-1} d_K f(y) + o\left(\frac{1}{Nh}\right). \tag{4.57}$$

Proof: First, we write

$$\begin{aligned}\sqrt{nh}[\hat{f}_d(y;h) - E_\xi\{f_U(y;h)\}] &= \sqrt{nh}[\hat{f}_d(y;h) - f_U(y;h)] \\ &\quad + \sqrt{nh}[f_U(y;h) - E_\xi\{f_U(y;h)\}]. \end{aligned} \tag{4.58}$$

From Lemma 2.4 and the assumption that $n_\tau/N_\tau \to \pi$ as $\tau \to \infty$, we have

$$\sqrt{nh}[f_U(y;h) - E_\xi\{f_U(y;h)\}] \xrightarrow{\mathscr{L}_\xi} N(0, \pi f(y) d_K). \tag{4.59}$$

Moreover, from (4.48),

$$[\hat{f}_d(y;h) - f_U(y;h)]\Lambda_{\mathcal{P}}^{-1/2} \xrightarrow{\mathscr{L}_{\mathcal{P}}} N(0,1), \tag{4.60}$$

where $\Lambda_{\mathcal{P}}$ is as in (4.42). Note that, from (4.42) and (4.50), $\Lambda_{\mathcal{P}} = (1 - n/N)[N/(N - 1)]\gamma_{\mathcal{P}}^2/nh^2$. Now,

$$\begin{aligned} E_\xi\left[\frac{1}{nh^2}\gamma_{\mathcal{P}}^2\right] &= \frac{1}{nNh^2}\sum_{i\in U} E_\xi\left[(u_i^*(h) - m_{Ui}(h)) - N^{-1}\{\sum_{i\in U}(u_i^*(h) - m_{Ui}(h))\}\right]^2 \\ &= \frac{1}{nNh^2}\left[\sum_{i\in U} E_\xi(u_i^*(h) - m_{Ui}(h))^2 - N^{-1}E_\xi\{\sum_{i\in U}(u_i^*(h) - m_{Ui}(h))\}^2\right] \\ &:= \Pi_1 - \Pi_2. \end{aligned} \tag{4.61}$$

Calculating each term separately,

$$
\begin{aligned}
\Pi_1 &= \frac{1}{nNh^2}\sum_{i\in U} E_\xi (u_i^*(h) - m_{Ui}(h))^2 \\
&= \frac{1}{nNh^2}\sum_{i\in U} E_\xi \left[\{u_i^*(h) - m(X_i;y)\} + \{m(X_i;y) - m_U(X_i;y)\}\right]^2 \\
&= \frac{1}{N}\sum_{i\in U}\left\{\frac{1}{nh^2}E_\xi \{u_i^*(h) - m(X_i;y)\}^2 + \frac{1}{nh^2}E_\xi \{m(X_i;y) - m_U(X_i;y)\}^2\right. \\
&\qquad\qquad \left. + \frac{2}{nh^2}E_\xi \left[\{u_i^*(h) - m(X_i;y)\}\{m(X_i;y) - m_U(X_i;y)\}\right]\right\} \\
&:= \frac{1}{N}\sum_{i\in U}\left[\Pi_{11} + \Pi_{12} + \Pi_{13}\right].
\end{aligned}
$$

From Eq. (4.22) in the proof of Theorem 4.2,

$$
\Pi_{11} = \frac{1}{nh} d_K f(y) + o\left(\frac{1}{nh}\right).
$$

Additionally, from Eq. (4.28),

$$
\Pi_{12} = o\left(b^4 + \frac{1}{nb}\right).
$$

Regarding Π_{13}, note that

$$
\begin{aligned}
\Pi_{13} = \frac{2}{nh^2}\big[&E_\xi\{u_i^*(h)m(X_i;y)\} - E_\xi\{m^2(X_i;y)\} + E_\xi\{m(X_i;y)m_U(X_i;y)\} \\
&- E_\xi\{u_i^*(h)m_U(X_i;y)\}\big].
\end{aligned} \tag{4.62}
$$

By calculating each term in (4.62), we can show that $\Pi_{13} = o(\{nh\}^{-1})$. Consequently,

$$
\Pi_1 = \frac{1}{nh} d_K f(y) + o\left(\frac{1}{nh}\right).
$$

Similarly,

$$
\Pi_2 = \frac{1}{nN^2h^2}E_\xi\left\{\sum_{i\in U}(u_i^*(h) - m_{Ui}(h))\right\}^2
$$

$$= \frac{1}{nN^2h^2} \sum_{i \in U} E_\xi (u_i^*(h) - m_{Ui}(h))^2$$
$$+ \frac{1}{nN^2h^2} \sum\sum_{i \neq j, i, j \in U} E_\xi (u_i^*(h) - m_{Ui}(h))(u_j^*(h) - m_{Uj}(h)) = o\left(\frac{1}{nh}\right).$$

Therefore,

$$E_\xi \left[\frac{1}{nh^2} \gamma_{\mathcal{P}}^2\right] = \frac{1}{nh} d_K f(y) + o\left(\frac{1}{nh}\right),$$

and, hence,

$$E_\xi(\Lambda_{\mathcal{P}}) = \left(1 - \frac{n}{N}\right) \frac{N}{N-1} \frac{1}{nh} d_K f(y) + o\left(\frac{1}{nh}\right). \tag{4.63}$$

From (4.59), (4.60) and (4.63), the conditions for Theorem 1.3.6 of Fuller (2009) are satisfied and, thus,

$$\sqrt{nh}\left[\hat{f}_d(y;h) - E_\xi\{f_U(y;h)\}\right] \xrightarrow{\mathcal{L}_C} N(0, V),$$

where

$$V = (1 - \pi) d_K f(y) + \pi d_K f(y) = d_K f(y).$$

The proof is complete upon noting that $V_C\{\hat{f}_d(y;h)\} = (nh)^{-1}\{V + o(1)\}$. $\square$

4.5. Bandwidth Selection

The estimator $\hat{f}_d(y;h)$ in (4.7) uses two smoothing parameters; the first of which (b) is used to smooth the regression estimator $\hat{m}(x; y, 1, b)$ (see Eq. (4.6)) while the second (h) determines the amount of smoothness of the density estimator itself. In this section, we discuss how these parameters can be chosen such that a good balance between the bias and the variance of $\hat{f}_d(y;h)$ is reached.

Starting with the bandwidth parameter for the density estimator itself, note that the asymptotic MISE of $\hat{f}_d(y;h)$ under the combined mode of inference is given by (see Theorem 4.2)

$$AMISE\left[\hat{f}_d(\cdot;h)\right] = (nh)^{-1}\delta^* d_K + \frac{1}{4}h^4 c_K^2 d_{f''}. \tag{4.64}$$

Using this formula for the $AMISE$, the asymptotically optimal bandwidth $h_{opt,d}$, in the sense that it minimizes the $AMISE$, follows:

$$h_{opt,d} = \left[\frac{\delta^* d_K}{c_K^2 d_{f''}}\right]^{1/5} n^{-1/5}. \tag{4.65}$$

Notice that the optimal bandwidth $h_{opt,d}$ in (4.65) is $O(n^{-1/5})$ which is identical to the bandwidth order in standard kernel density estimators with no auxiliary data. The optimal bandwidth for the estimator $\hat{f}_d(y;h)$ is adjusted by the sampling design effect represented by δ^*. Since this bandwidth is chosen to minimize the asymptotic MISE, it does not show the effect of auxiliary data which is embedded in the small-order terms of the MISE formula as shown in (4.17). Also, observe that $h_{opt,d}$ is unobtainable because it involves the density functional $d_{f''}$. A simple way to obtain an estimate for $h_{opt,d}$ is to replace the functional $d_{f''}$ by a kernel estimate. This is the plug-in approach we described in Section 2.5. Following the four steps in Section 2.5, we obtain the following plug-in estimate for $h_{opt,d}$:

$$\hat{h}_{DPI,d} = \left[\frac{\delta^* d_K}{c_K^2 \hat{\Psi}_4(h^*)}\right]^{1/5} n^{-1/5}, \tag{4.66}$$

where $\hat{\Psi}_r(h^*)$ is as in (2.70) with h^* replacing b_1 (see Section 2.5).

Now, it remains to choose the smoothing parameter b for the local linear regression estimator $\hat{m}(x;y,1,b)$. There exist a fair amount of literature on the problem of selecting smoothing parameters for local linear regression estimates. To mention a few, we refer to Fan and Gijbels (1995), Ruppert et al. (1995) and Hengartner et al. (2002). Ruppert et al. (1995) developed plug-in estimators for the bandwidth of local linear regression estimators. Such estimators were shown to have a very reliable performance both theoretically and practically. As we discussed in different places of the preceding chapters, plug-in bandwidth estimators are built by replacing the unknown functionals in the formula of the asymptotically optimal bandwidth by kernel estimators. Thus, to define a plug-in estimate for b, we need to define the asymptotically optimal b first. That is, we need to find b that minimizes the AMISE of $\hat{m}(x;y,1,b)$. Following an argument similar to that in (Wand and Jones (1995, pg.123-124)), we can show that the conditional MSE of $\hat{m}(x;y,1,b)$ is

$$\begin{aligned} & MSE_{\xi}\left[\hat{m}(x;y,1,b)|X_1,\ldots,X_N\right] \\ & \qquad = \frac{1}{4}c_{\psi}^2\{m''(x;y)\}^2 b^4 + \frac{1}{nb}\kappa d_{\psi}\frac{\sigma^2(x;y)}{g(x)} + o_p\left(b^4+\frac{1}{Nb}\right), \end{aligned} \tag{4.67}$$

where $m''(x;y) = (\partial^2/\partial x^2)m(x;y)$ and $\kappa = nN^{-1}\{N^{-1}\sum_{i\in s} d_i^2\}\{(N^{-1}\sum_{i\in s} d_i)^2\}^{-1}$ is a factor that reflects the effect of incorporating the design weights into $\hat{m}(x;y,1,b)$. For SRSWOR, $d_i = Nn^{-1}$ for all $i \in s$ and, thus, $\kappa = 1$. Multiplying each of the first two terms in (4.67) by $g(x)$ and integrating over x gives the following weighted conditional AMISE:

$$AWMISE\left[\hat{m}(.;y,1,b)\right] = \frac{1}{4}c_{\psi}^2 b^4 \int \{m''(x;y)\}^2 g(x)dx + \frac{1}{nb}\kappa d_{\psi}\int \sigma^2(x)dx. \tag{4.68}$$

Now using (4.68), we can derive the asymptotically optimal bandwidth for the regression estimate $\hat{m}(x;y,1,b)$:

$$b_{opt} = \left[\frac{d_{\psi}\kappa\int\sigma^2(x)dx}{nc_{\psi}^2\theta_{22}}\right]^{1/5}, \tag{4.69}$$

where $\theta_{rs} = \int\{m^{(r)}(x;y)m^{(s)}(x;y)\}g(x)dx$. Assuming that X has the bounded support $[0,1]$ and the model is homescedastic with $\sigma^2(x;y) = \sigma^2$, we can simplify (4.69) to get:

$$b_{opt} = \left[\frac{d_{\psi}\kappa\sigma^2}{nc_{\psi}^2\theta_{22}}\right]^{1/5}. \tag{4.70}$$

A direct plug-in estimate of b_{opt} is then obtained by replacing the unknown quantities σ^2

and θ_{22} in (4.70) by kernel estimates (e.g., Ruppert et al. (1995)):

$$\hat{b}_{DPI} = \left[\frac{d_\psi \kappa \hat{\sigma}^2}{n c_\psi^2 \hat{\theta}_{22}}\right]^{1/5}, \tag{4.71}$$

where

$$\hat{\sigma}^2(\lambda) = \{n - 2\sum_{i=1}^{n} w_{ii} + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^2\}^{-1} \sum_{i=1}^{n}\{Y_i - \hat{m}(X_i;1,\lambda)\}^2, \tag{4.72}$$

and

$$\hat{\theta}_{22}(\varsigma) = n^{-1}\sum_{i=1}^{n} \widehat{m''}(X_i;3,\varsigma), \tag{4.73}$$

$w_{ij} = \mathbf{e}_1^T \left(\mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{X}_{is}\right)^{-1} \mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{e}_j$, $\hat{m}(X_i;1,\lambda) = \mathbf{e}_1^T \left(\mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{X}_{is}\right)^{-1} \mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{z}_s$, $\widehat{m''}(X_i;3,\varsigma) = 2\mathbf{e}_3^T \left(\mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{X}_{is}\right)^{-1} \mathbf{X}_{is}^T \mathbf{W}_{is} \mathbf{z}_s$, $\mathbf{W}_{is} = diag\left\{\psi_\lambda(x_i - X_j)\right\}_{j \in s}$, $\mathbf{X}_{is}$ and $\mathbf{z}_s$ have the same form as in Section 4.2 and λ and ς are two new bandwidth parameters which will depend on σ^2 and θ_{24}. For simplicity, these new bandwidths are commonly estimated using quick and simple methods such as the blocking method described in Ruppert et al. (1995).

4.6. Proofs of Technical Lemmas

Proof of Lemma 4.1: Note that $m_U(x;y,1,b)$, given in (4.5), is the standard local linear regression estimator for $m(x;y)$ in model (4.2). This estimator uses all N observations in the finite population. In model (4.2), the response variable is $Z_i = K\left(h^{-1}(Y_i - y)\right)$ which is always bounded due to the boundedness of the kernel function K under assumption A.2(i). Thus, given conditions C.1(ii) and C.2, the proof of the lemma follows directly from standard results on local linear regression (e.g., Wand and Jones (1995, pg. 124)). □

Proof of Lemma 4.2: First, rewrite the local linear fits in (4.5) and (4.6) as follows:

$$m_U(x_i;y) = \frac{S_{i3}S_{i4} - S_{i2}S_{i5}}{S_{i1}S_{i3} - S_{i2}^2} \tag{4.74}$$

and

$$\hat{m}(x_i;y) = \frac{\hat{S}_{i3}\hat{S}_{i4} - \hat{S}_{i2}\hat{S}_{i5}}{\hat{S}_{i1}\hat{S}_{i3} - \hat{S}_{i2}^2}, \tag{4.75}$$

where $S_{ij} = N^{-1}\sum_{k\in U} V_{ijk}$, $\hat{S}_{ij} = N^{-1}\sum_{k\in U} I_k d_k V_{ijk}$ and

$$V_{ijk} = \begin{cases} \psi_b(x_k - x_i)(x_k - x_i)^{(j-1)}, & j = 1,2,3, \\ \psi_b(x_k - x_i)(x_k - x_i)^{(j-4)} z_k, & j = 4,5. \end{cases}$$

Setting $\mathbf{S}_i = (S_{i1},\ldots,S_{i5})^T$ and $\hat{\mathbf{S}}_i = (\hat{S}_{i1},\ldots,\hat{S}_{i5})^T$, we can write $m_U(x_i;y) = l(\mathbf{S}_i)$ and $\hat{m}(x_i;y) = l(\hat{\mathbf{S}}_i)$. Using a multivariate version of Taylor Theorem, we have

$$\hat{m}(x_i;y) = m_U(x_i;y) + \sum_{j=1}^{5} \left(\left.\frac{\partial \hat{m}(x_i;y)}{\partial \hat{S}_{ij}}\right|_{\hat{\mathbf{S}}_i = \mathbf{S}_i} \right) (\hat{S}_{ij} - S_{ij}) + R_i,$$

where R_i is the remainder. This linearization step can be rewritten as

$$\hat{m}(x_i;y) - m_U(x_i;y) = \sum_{j=1}^{5} \left.\frac{\partial \hat{m}(x_i;y)}{\partial \hat{S}_{ij}}\right|_{\hat{\mathbf{S}}_i = \mathbf{S}_i} (\hat{S}_{ij} - S_{ij}) + R_i.$$

It is not hard to see that

$$\hat{S}_{ij} - S_{ij} = \frac{1}{N}\sum_{k\in U} (I_k d_k - 1) V_{ijk},$$

which implies that

$$\hat{m}(x_i;y) - m_U(x_i;y) = \frac{1}{N}\sum_{k\in U} (I_k d_k - 1) V_{ik}^* + R_i, \tag{4.76}$$

where

$$V_{ik}^* = \sum_{j=1}^{5} \left.\frac{\partial \hat{m}(x_i;y)}{\partial \hat{S}_{ij}}\right|_{\hat{\mathbf{S}}_i = \mathbf{S}_i} V_{ijk}.$$

Therefore, we can write

$$
\begin{aligned}
&(\hat{m}(x_i;y)-m_U(x_i;y))(\hat{m}(x_j;y)-m_U(x_j;y))(I_id_i-1)(I_jd_j-1)\\
&\quad=\left(\frac{1}{N}\sum_{k\in U}(I_kd_k-1)V_{ik}^*+R_i\right)\left(\frac{1}{N}\sum_{k\in U}(I_kd_k-1)V_{jk}^*+R_j\right)\\
&\quad=\frac{1}{N^2}\sum\sum_{k,l\in U}(I_kd_k-1)(I_ld_l-1)V_{ik}^*V_{il}^*+\frac{2}{N}R_j\sum_{k\in U}(I_kd_k-1)V_{ik}^*+R_iR_j,
\end{aligned}
$$

and, consequently,

$$
\begin{aligned}
\Upsilon &:= \frac{1}{N^2h^2}\sum\sum_{i,j\in U}(\hat{m}(X_i;y)-m_U(X_i;y))(\hat{m}(X_j;y)-m_U(X_j;y))(I_id_i-1)(I_jd_j-1)\\
&= \frac{1}{N^4h^2}\sum\cdots\sum_{i,j,k,l\in U}V_{ik}^*V_{il}^*(I_id_i-1)(I_jd_j-1)(I_kd_k-1)(I_ld_l-1)\\
&\quad+\frac{2}{N^3h^2}\sum\sum\sum_{i,j,k\in U}V_{ik}^*R_j(I_id_i-1)(I_jd_j-1)(I_kd_k-1)\\
&\quad+\frac{1}{N^2h^2}\sum\sum_{i,j\in U}R_iR_j(I_id_i-1)(I_jd_j-1)\\
&:= \Upsilon_1+\Upsilon_2+\Upsilon_3.
\end{aligned}
\tag{4.77}
$$

Now, observe that

$$
\begin{aligned}
E_{\mathcal{P}}(\Upsilon_1) \;=\; &\frac{1}{N^4h^2}\sum_{i\in U}V_{ii}^{*2}E_{\mathcal{P}}\{(I_id_i-1)^4\}+\frac{1}{N^4h^2}\sum\sum_{i,j\in U,i\neq j}V_{ij}^{*2}E_{\mathcal{P}}\{(I_id_i-1)^3(I_jd_j-1)\}\\
&+\frac{2}{N^4h^2}\sum\sum_{i,j\in U,i\neq j}V_{ii}^*V_{ij}^*E_{\mathcal{P}}\{(I_id_i-1)^3(I_jd_j-1)\}\\
&+\frac{1}{N^4h^2}\sum\sum_{i,j\in U,i\neq j}V_{ii}^{*2}E_{\mathcal{P}}\{(I_id_i-1)^3(I_jd_j-1)\}\\
&+\frac{1}{N^4h^2}\sum\sum_{i,j\in U,i\neq j}V_{ij}^{*2}E_{\mathcal{P}}\{(I_id_i-1)^2(I_jd_j-1)^2\}\\
&+\frac{2}{N^4h^2}\sum\sum_{i,j\in U,i\neq j}V_{ii}^*V_{ij}^*E_{\mathcal{P}}\{(I_id_i-1)^2(I_jd_j-1)^2\}\\
&+\frac{3}{N^4h^2}\sum\sum\sum_{i,j,k\in U,i\neq j\neq k}V_{ij}^*V_{ik}^*E_{\mathcal{P}}\{(I_id_i-1)^2(I_jd_j-1)(I_kd_k-1)\}\\
&+\frac{2}{N^4h^2}\sum\sum\sum_{i,j,k\in U,i\neq j\neq k}V_{ii}^*V_{ij}^*E_{\mathcal{P}}\{(I_id_i-1)^2(I_jd_j-1)(I_kd_k-1)\}\\
&+\frac{1}{N^4h^2}\sum\sum\sum_{i,j,k\in U,i\neq j\neq k}V_{ij}^{*2}E_{\mathcal{P}}\{(I_id_i-1)^2(I_jd_j-1)(I_kd_k-1)\}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{N^4h^2}\sum_{i,j,k,l\in U, i\neq j\neq k\neq l}\cdots\sum V_{ik}^{*}V_{il}^{*}E_{\mathcal{P}}\{(I_i d_i-1)(I_j d_j-1)(I_k d_k-1)(I_l d_l-1)\}\\
&:=\Upsilon_{1,1}+\Upsilon_{1,2}+\cdots+\Upsilon_{1,10}.
\end{aligned}
\tag{4.78}
$$

Recall that the indicator variables I_i, defined in the proof of Lemma 2.2 in Section 2.3, are Bernoulli random variables with mean π_i and that $d_i = \pi_i^{-1}$, for all $i \in U$.

$$
\begin{aligned}
\Upsilon_{1,10} &= \frac{1}{N^4h^2}\sum_{\substack{i,j,k,l\in U,\\ i\neq j\neq k\neq l}}\cdots\sum V_{ik}^{*}V_{il}^{*}\frac{E_{\mathcal{P}}\{(I_i-\pi_i)(I_j-\pi_j)(I_k-\pi_k)(I_l-\pi_l)\}}{\pi_i\pi_j\pi_k\pi_l}\\
&\le \frac{\lambda^{-4}}{n^2N^4h^2}\sum_{\substack{i,j,k,l\in U,\\ i\neq j\neq k\neq l}}\cdots\sum V_{ik}^{*}V_{il}^{*}n^2\max_{\substack{i,j,k,l\in U,\\ i\neq j\neq k\neq l}}|E_{\mathcal{P}}\{(I_i-\pi_i)(I_j-\pi_j)(I_k-\pi_k)(I_l-\pi_l)\}|\\
&= O_p\left(\frac{1}{n^2h^2}\right),
\end{aligned}
$$

where the last equality follows from Assumption C.3(i). Additionally,

$$
\begin{aligned}
\Upsilon_{1,9} &= \frac{1}{N^4h^2}\sum\sum\sum_{i,j,k\in U, i\neq j\neq k} V_{ij}^{*2}\frac{E_{\mathcal{P}}\{(I_i-\pi_i)^2(I_j-\pi_j)(I_k-\pi_k)\}}{\pi_i^2\pi_j\pi_k}\\
&\le \frac{\lambda^{-4}}{nN^4h^2}\sum\sum\sum_{\substack{i,j,k\in U,\\ i\neq j\neq k}} V_{ij}^{*2}n\max_{\substack{i,j,k\in U,\\ i\neq j\neq k}}|E_{\mathcal{P}}\{(I_i-\pi_i)^2(I_j-\pi_j)(I_k-\pi_k)\}| = O_p\left(\frac{1}{nNh^2}\right),
\end{aligned}
$$

where the last equality follows from Assumption C.3(ii). Similar bounding arguments give the same bound for $\Upsilon_{1,8}$ and $\Upsilon_{1,7}$.

$$
\begin{aligned}
\Upsilon_{1,6} &= \frac{2}{N^4h^2}\sum\sum_{i,j\in U, i\neq j} V_{ii}^{*}V_{ij}^{*}\frac{E_{\mathcal{P}}\{(I_i-\pi_i)^2(I_j-\pi_j)^2\}}{\pi_i^2\pi_j^2}\\
&\le \frac{2\lambda^{-4}}{N^4h^2}\sum\sum_{i,j\in U, i\neq j} V_{ii}^{*}V_{ij}^{*}\left[E_{\mathcal{P}}\{(I_i-\pi_i)^4\}E_{\mathcal{P}}\{(I_i-\pi_i)^4\}\right]^{1/2} = O_p\left(\frac{1}{N^2h^2}\right),
\end{aligned}
$$

where the we used the Cauchy-Schwarz inequality in the second line and the last equality follows from the boundedness of the moments of the Bernoulli distribution. Similarly, $\Upsilon_{1,5}$,

$\Upsilon_{1,4}$, $\Upsilon_{1,3}$ and $\Upsilon_{1,2}$ have the same bound as $\Upsilon_{1,6}$. Finally,

$$\begin{aligned}\Upsilon_{1,1} &= \frac{1}{N^4h^2}\sum_{i\in U} V_{ii}^{*2} E_{\mathcal{P}}\{(I_i-\pi_i)^4\}\pi_i^{-4}\\ &= \frac{1}{N^4h^2}\sum_{i\in U} V_{ii}^{*2}\pi_i(1-\pi_i)(3\pi_i^2-3\pi_i+1)\pi_i^{-4}\\ &\leq \frac{1}{N^4h^2}\sum_{i\in U} V_{ii}^{*2}\pi_i(1-\pi_i)(3\pi_i^2-3\pi_i+1)\lambda^{-4} = O_p\left(\frac{1}{N^3h^3}\right),\end{aligned}$$

where the inequality follows from Assumption A.4. Substituting these results into (4.78), gives

$$E_{\mathcal{P}}(\Upsilon_1) = o_p\left(\frac{1}{Nh}\right). \tag{4.79}$$

On the other hand,

$$\begin{aligned}E_{\mathcal{P}}(\Upsilon_3) &= \frac{1}{N^2h^2}\sum_{i\in U} E_{\mathcal{P}}\{R_i^2(I_id_i-1)^2\} + \frac{1}{N^2h^2}\sum\sum_{i,j\in U, i\neq j} E_{\mathcal{P}}\{R_iR_j(I_id_i-1)(I_jd_j-1)\}\\ &= \Upsilon_{3,1}+\Upsilon_{3,2}. \end{aligned}\tag{4.80}$$

$$\begin{aligned}\Upsilon_{3,1} &= \frac{1}{N^2h^2}\sum_{i\in U} E_{\mathcal{P}}\{R_i^2(I_id_i-1)^2\}\\ &= \frac{1}{N^2h^2}\sum_{i\in U} E_{\mathcal{P}}(R_i^2)E_{\mathcal{P}}\{(I_i-\pi_i)^2\}\pi_i^{-2}\\ &= \frac{1}{N^2h^2}\sum_{i\in U} E_{\mathcal{P}}(R_i^2)\frac{\pi_i(1-\pi_i)}{\pi_i^2}\\ &\leq \frac{1}{N^2h^2}\sum_{i\in U} E_{\mathcal{P}}(R_i^2)\frac{(1-\lambda)}{\lambda} = O_p\left(\frac{1}{n^2Nh^4}\right),\end{aligned}$$

where the second equality follows because R_i is independent of the sample indicators, the inequality in the fourth line follows from Assumption A.4 and the last equality follows from the fact that $\frac{n}{N}\sum_{i\in U} E_{\mathcal{P}}(R_i^2) = O_p\left(\frac{1}{nh^2}\right)$ (see Lemma 3 in Breidt and Opsomer (2000)).

$$\Upsilon_{3,2} = \frac{1}{N^2h^2}\sum\sum_{i,j\in U, i\neq j} E_{\mathcal{P}}(R_iR_j)E_{\mathcal{P}}\{(I_i-\pi_i)(I_j-\pi_j)\}\pi_i\pi_j$$

$$
\begin{aligned}
&= \frac{1}{N^2h^2} \sum\sum_{i,j\in U, i\neq j} E_{\mathcal{P}}(R_iR_j) \frac{(\pi_{ij} - \pi_i\pi_j)}{\pi_i\pi_j} \\
&\leq \frac{1}{nN^2h^2} \sum\sum_{i,j\in U, i\neq j} E_{\mathcal{P}}(R_iR_j) n \max_{i,j\in U, i\neq j} |\pi_{ij} - \pi_i\pi_j| \lambda^{-2} \\
&\leq \frac{1}{nNh^2} \sum_{i\in U} E_{\mathcal{P}}(R_i^2) n \max_{i,j\in U, i\neq j} |\pi_{ij} - \pi_i\pi_j| \lambda^{-2} \\
&= \frac{1}{n^2h^2} \left[\frac{n}{N} \sum_{i\in U} E_{\mathcal{P}}(R_i^2) \right] n \max_{i,j\in U, i\neq j} |\pi_{ij} - \pi_i\pi_j| \lambda^{-2} = O_p\left(\frac{1}{n^3h^4}\right),
\end{aligned}
$$

where the last equality follows from the fact that $n \max_{i,j\in U, i\neq j} |\pi_{ij} - \pi_i\pi_j| = O(1)$ by Assumption A.4. Substituting these results into (4.80), we have

$$
E_{\mathcal{P}}(\Upsilon_3) = o_p\left(\frac{1}{nh}\right). \tag{4.81}
$$

Finally, notice that

$$
\Upsilon_2 = 2\Upsilon_1^{1/2}\Upsilon_3^{1/2}.
$$

Therefore, an application of the Cauchy-Schwarz inequality shows that

$$
\Upsilon_2 = o_p\left(\frac{1}{nh}\right). \tag{4.82}
$$

Using (4.79), (4.81) and (4.82) together with (4.77) gives the result. □

CHAPTER 5

A Simulation Study

In this chapter, we investigate the finite sample properties of the three proposed model-assisted kernel density estimators of Chapters 2–4, namely, $\hat{f}_{dl}(\cdot)$, $\hat{f}_{dn}(\cdot)$ and $\hat{f}_{d}(\cdot)$, through a Monte Carlo study. We consider comparing the performance of these estimators to that of the weighted kernel density estimator $\hat{f}_{w}(\cdot)$ given in (1.12), which ignores the auxiliary information. This comparison is held under finite populations drawn from different distributions. We also consider several forms for the relationship between the study variable and the auxiliary variable. Both equal-probability and unequal-probability sampling plans are used. In the following section, we give the details of the different settings considered in this simulation study. The simulation results are reported in Section 5.2.

5.1. Simulation Settings

We consider estimating the density function of study variables having three distributions belonging to three different distribution families:

- The standard normal distribution $[N(0,1)]$;
- The mixture normal distribution $[0.5N(-1,2/3)+0.5N(1,2/3)]$;
- The skew normal distribution $[ESN(-0.5,1.5,10,1)]$.

The density functions of these distributions are displayed in Figure 5.1. In our simulation, we start by generating the finite population values for the study variable, Y, from each of the three distributions mentioned above (we use finite populations of size $N=1,000$). Then, we generate the finite population values for the auxiliary variable, X, such that the relationship between X and Y takes one of the following three forms:

- Model I. $Y = 1 + 2*(X - 0.5) + \varepsilon$;
- Model II. $Y = 2.5*ln(X + 1.5) + \varepsilon$;
- Model III. $Y = \pm\sqrt{-2*ln(X)} + 0.6\varepsilon$,

where $\varepsilon \sim N(0, \sigma^2)$. The three values $\sigma = 0.2, 0.6, 1.0$ are used to account for three correlation levels; strong, moderate and weak, respectively.

We use both SRSWOR and Poisson sampling to sample from the finite populations. The first sampling plan is an equal-probability sampling plan while the second is an unequal-probability sampling scheme. For Poisson sampling, the sampling weights are defined such that they are proportional to the design variable $w_i = \sqrt{x_i + c}$ for $i \in U$ where c is a constant that makes the quantity under the square root positive and it is set to be $c = 3.5, 1.5, 0.01$ for models I, II and III, respectively. Poisson sampling is implemented using the R (R Development Core Team (2011)) function *UPpoisson* inside the package *sampling*. The function *UPpoisson* standardizes the sampling weights d_i such that $\sum_{i \in U} d_i^{-1} = n$. Four sample sizes are considered; $n = 25, 50, 75, 100$.

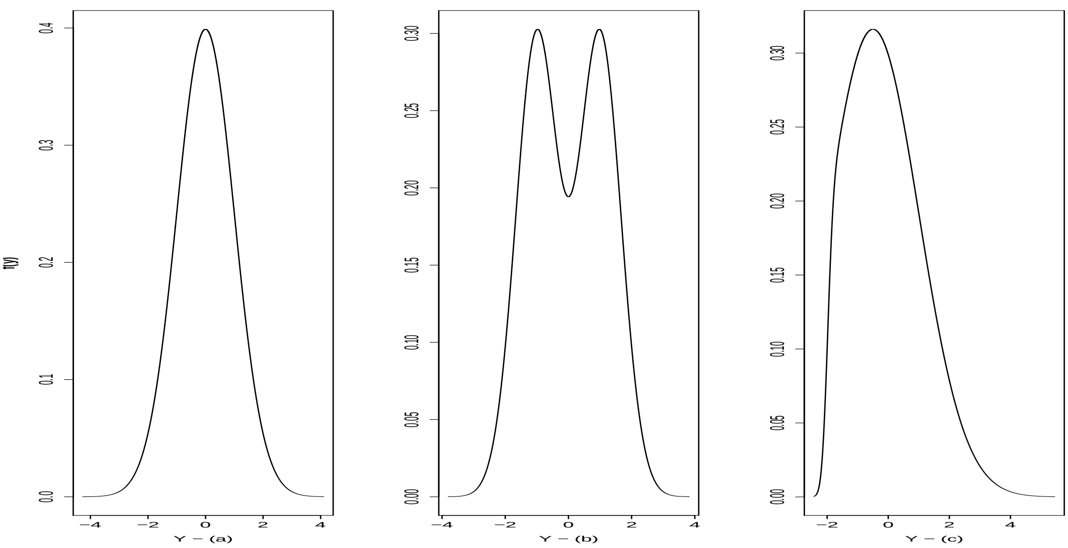

Figure 5.1: True densities for the study variable Y: (a) $N(0,1)$; (b) $0.5N(-1, 2/3) + 0.5N(1, 2/3)$ and (c) $ESN(-0.5, 1.5, 10, 1)$.

Under each of the above simulation settings, a single finite population of size $N = 1,000$ is generated first and then we draw repeated samples ($m = 1,000$ samples) from that finite population using either SRSWOR or Poisson sampling. The four estimators $\hat{f}_w(\cdot)$, $\hat{f}_{dl}(\cdot)$, $\hat{f}_{dn}(\cdot)$ and $\hat{f}_d(\cdot)$ are calculated from each sample at 201 grid points covering the range of Y. The standard estimator $\hat{f}_w(\cdot)$ is calculated using the plug-in bandwidth due to Sheather and Jones (1991) which is implemented inside the R function *density* in the package *stats*. In case of Poisson sampling, this bandwidth is adjusted by a factor to reflect the design-effect. The other three estimators, $\hat{f}_{dl}(\cdot)$, $\hat{f}_{dn}(\cdot)$ and $\hat{f}_d(\cdot)$, are calculated using the plug-in bandwidth estimates described in Sections 2.5, 3.5 and 4.5, respectively. The standard normal kernel is used in computing all density and regression estimators. All computations are implemented inside the R package.

Four performance measures, namely, bias, variance, MSE and MISE, are used to compare the performance of the four estimators listed above. The Monte Carlo versions of these measures are defined as follows:

$$BIAS_{MC}[\hat{f}(y)] = \frac{1}{m}\sum_{i=1}^{m} \hat{f}^i(y) - f(y), \tag{5.1}$$

$$V_{MC}[\hat{f}(y)] = \frac{1}{m}\sum_{i=1}^{m} \left[\hat{f}^i(y) - \left\{\frac{1}{m}\sum_{i=1}^{m} \hat{f}^i(y)\right\}\right]^2, \tag{5.2}$$

$$MSE_{MC}[\hat{f}(y)] = \frac{1}{m}\sum_{i=1}^{m} \left[\hat{f}^i(y) - f(y)\right]^2, \tag{5.3}$$

and

$$MISE_{MC}[\hat{f}(\cdot)] = \int MSE_{MC}[\hat{f}(y)]dy, \tag{5.4}$$

where the integral in (5.4) is evaluated using methods of numerical integration.

5.2. Simulation Results

The results of our simulation study are depicted in Figures 5.2–5.55 and Tables 5.1–5.6. Each of these figures displays the MSE curves of the four estimators, $\hat{f}_w(\cdot)$, $\hat{f}_{dl}(\cdot)$, $\hat{f}_{dn}(\cdot)$ and $\hat{f}_d(y)$, for the four sample sizes, $n = 25, 50, 75, 100$, under one of the simulation settings described in the previous section. Tables 5.1–5.6 report the MISE values of each estimator, i.e., the areas under the MSE curves for that estimator. Monte Carlo bias and variance were also calculated but not reported here because they give the same conclusions obtained from the MSE graphs. In the following, we discuss the main conclusions that come out of these figures and tables.

When the relationship between X and Y is linear, i.e., Model I holds true, it is evident that the three model-assisted density estimators $\hat{f}_{dl}(\cdot)$, $\hat{f}_{dn}(\cdot)$ and $\hat{f}_d(\cdot)$ are superior to the standard estimator $\hat{f}_w(\cdot)$ which ignores the auxiliary information. This conclusion holds for the three distributions under consideration, the two sampling plans, the four sample sizes and the three correlation levels (see the first section of Tables 5.1–5.6 and Figures 5.2–5.4, 5.11–5.13, 5.20–5.22 and their corresponding figures under Poisson sampling). This result should not be surprising since Model I favors the estimator $\hat{f}_{dl}(\cdot)$ and the other two estimators fit nonparametric models and, thus, they can detect the linear relationship. Under Model I, when the correlation between X and Y is weak, the four estimators perform closely for large samples ($n = 100$).

On the other hand, if Model II is the true model, i.e., the relationship between X and Y has a logarithmic pattern, the model-assisted estimator $\hat{f}_{dn}(\cdot)$ of Chapter 3 outperforms all other three estimators regardless of the shape of the density under estimation and the sampling plan (see the second section of Tables 5.1–5.6 and Figures 5.5–5.7, 5.14–5.16, 5.23–5.25 and their corresponding figures under Poisson sampling). In this case, the underlying model, Model II, favors $\hat{f}_{dn}(\cdot)$ while the regression function is considered misspecified for the estimators $\hat{f}_{dl}(\cdot)$ and $\hat{f}_d(\cdot)$. Despite this fact, the estimator $\hat{f}_{dl}(\cdot)$ continues to outperform the standard estimator $\hat{f}_w(\cdot)$ specially for weak to moderate correlations and small to moderate sample sizes (see Figures 5.6, 5.7, 5.16, 5.17, 5.26 and 5.27 and their corresponding figures under Poisson sampling).

Considering Model III, which favors the estimator $\hat{f}_d(\cdot)$ of Chapter 4, we find that $\hat{f}_d(\cdot)$ outperforms all other three estimators regardless of the shape of the density under estimation, the sampling plan, the sample size and the correlation level between X and Y (see the last section of Tables 5.1–5.6 and Figures 5.8–5.10, 5.17–5.19, 5.26–5.28 and their corresponding figures under Poisson sampling). Under this model, the other two model-assisted estimators $\hat{f}_{dl}(\cdot)$ and $\hat{f}_{dn}(\cdot)$ continue to perform better than the standard estimator $\hat{f}_w(\cdot)$ for small to medium sample sizes.

It is noteworthy that in all cases, as the sample size increases, the MISE of all four density estimators drops down confirming the consistency of these estimators. We should also note that the random sample size arising from Poisson sampling causes the MISE values of all four estimators to increase as compared to the MISE values under SRSWOR (compare the values in Tables 5.1–5.3 with those in Tables 5.4–5.6). A similar note can be made from the simulation study in Harms and Duchesne (2010) who studied nonparametric kernel regression estimation from complex survey data. Kernel density and regression estimations from samples with random sizes are covered in more details in the next chapter.

In general, it is clear from this simulation study that utilizing the available auxiliary information appropriately can significantly improve the efficiency of kernel density estimators. The three proposed model-assisted estimators $\hat{f}_{dl}(\cdot)$, $\hat{f}_{dn}(\cdot)$ and $\hat{f}_d(\cdot)$, although not meant to cover all possible cases, provide a good combination of kernel density estimators which utilize the available auxiliary information under a fair number of forms for the relationship between the study variable and the auxiliary one.

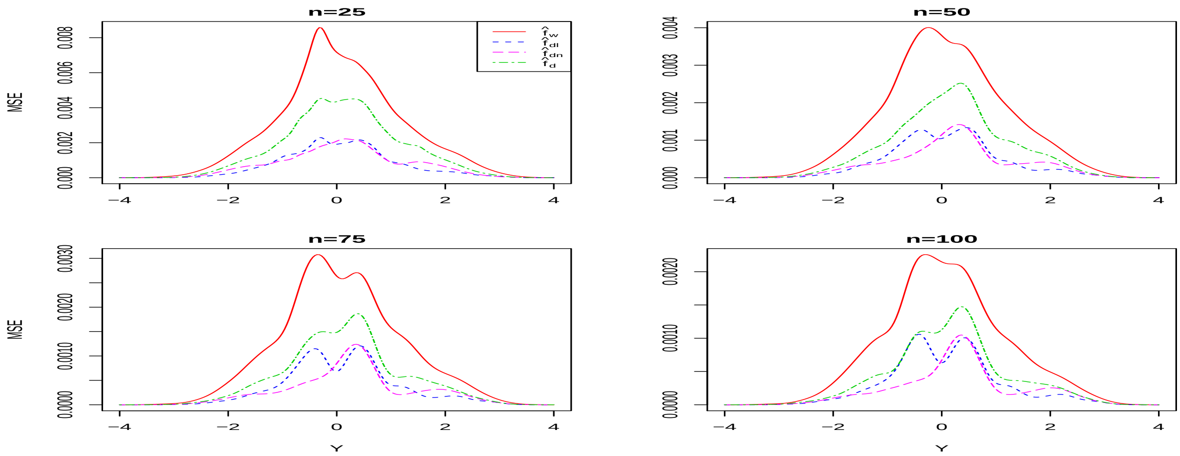

Figure 5.2: MSE for four KDEs under SRSWOR from the standard normal distribution - model I with $\sigma = 0.20$.

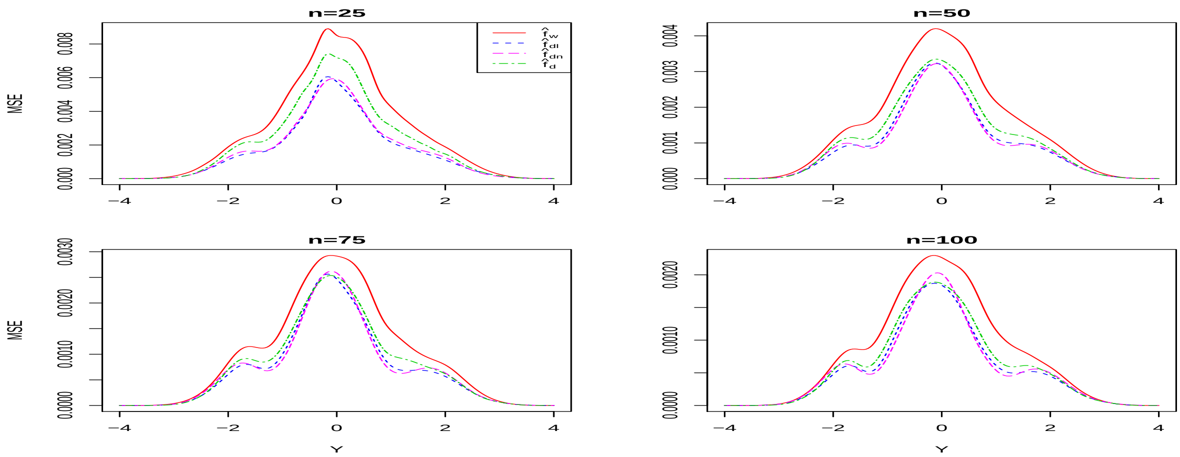

Figure 5.3: MSE for four KDEs under SRSWOR from the standard normal distribution - model I with $\sigma = 0.60$.

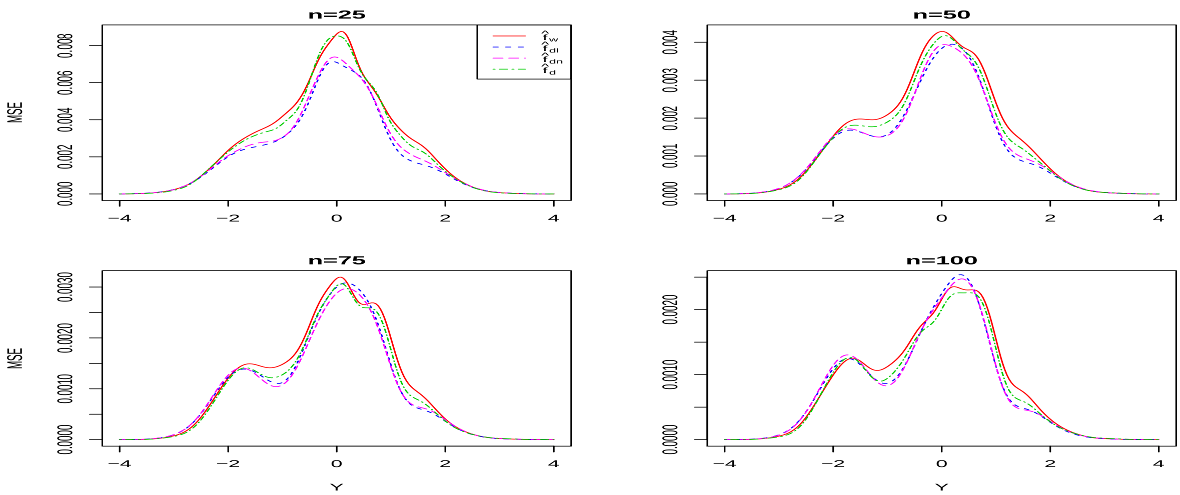

Figure 5.4: MSE for four KDEs under SRSWOR from the standard normal distribution - model I with $\sigma = 1.00$.

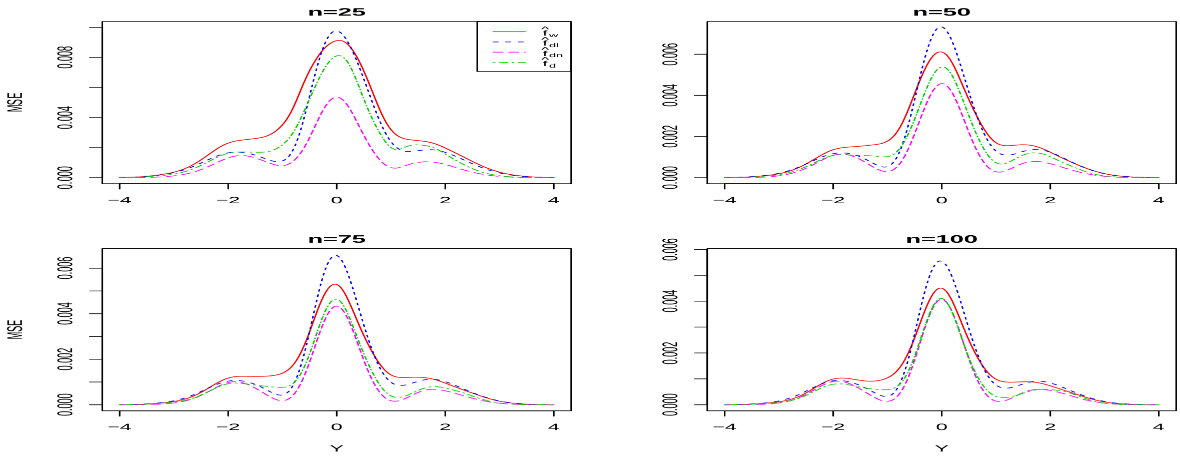

Figure 5.5: MSE for four KDEs under SRSWOR from the standard normal distribution - model II with $\sigma = 0.20$.

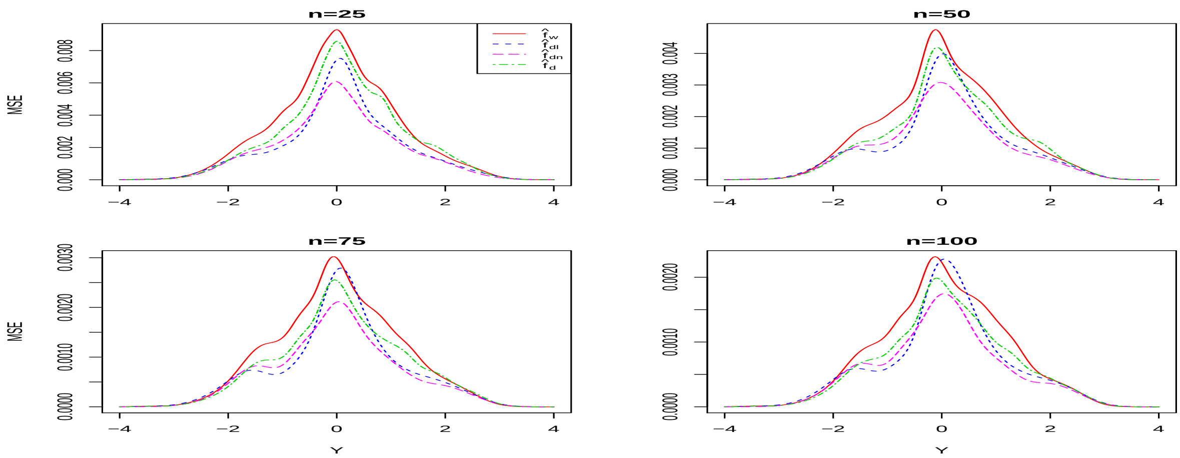

Figure 5.6: MSE for four KDEs under SRSWOR from the standard normal distribution - model II with $\sigma = 0.60$.

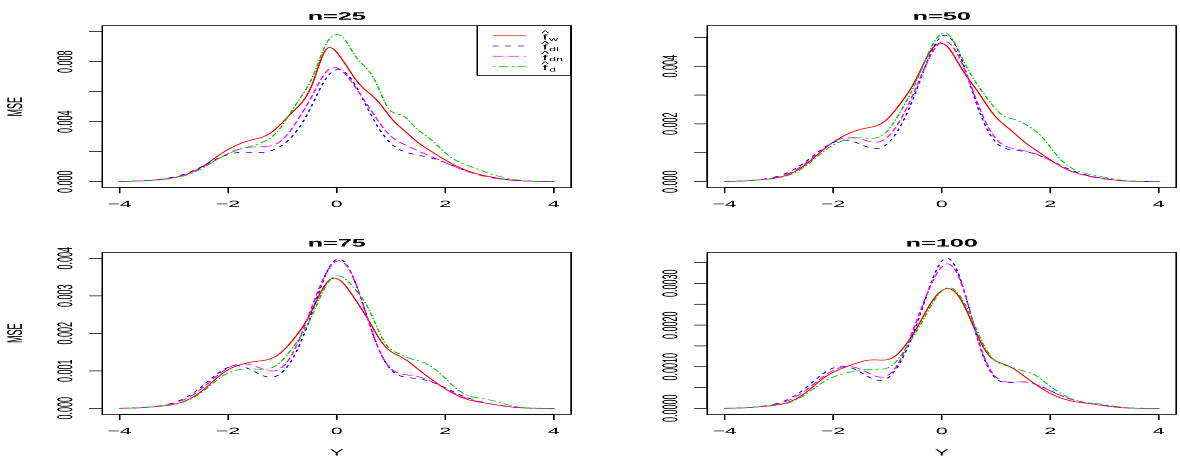

Figure 5.7: MSE for four KDEs under SRSWOR from the standard normal distribution - model II with $\sigma = 1.00$.

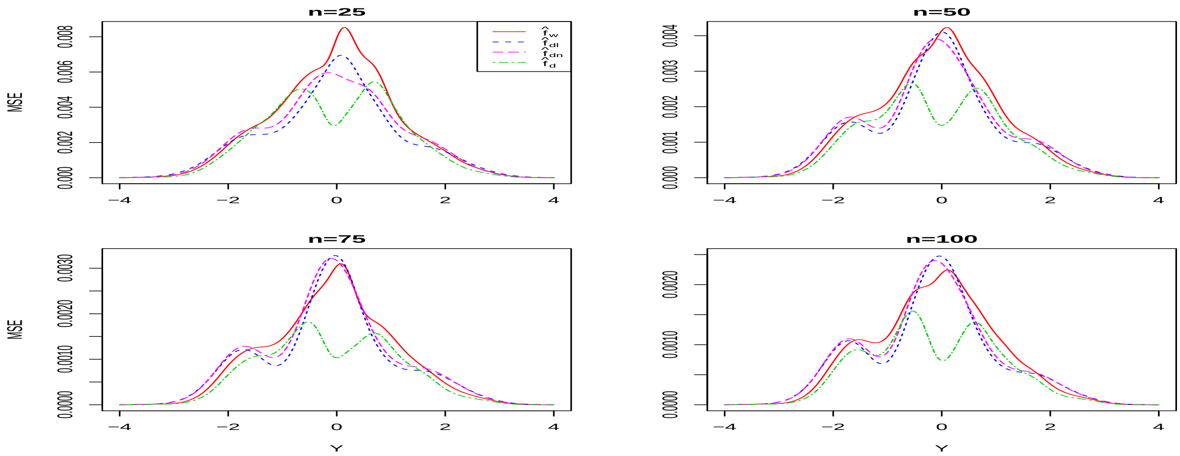

Figure 5.8: MSE for four KDEs under SRSWOR from the standard normal distribution - model III with $\sigma = 0.20$.

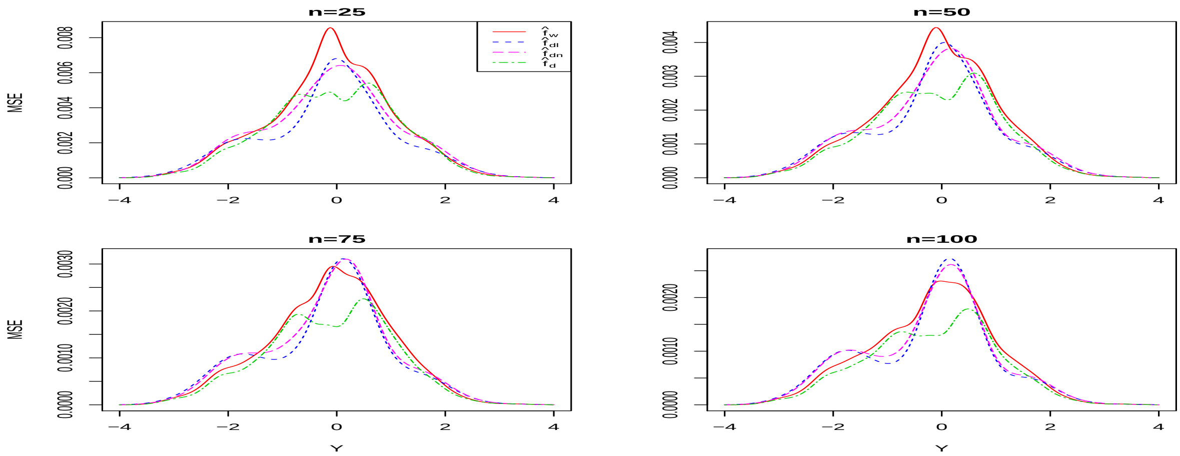

Figure 5.9: MSE for four KDEs under SRSWOR from the standard normal distribution - model III with $\sigma = 0.60$.

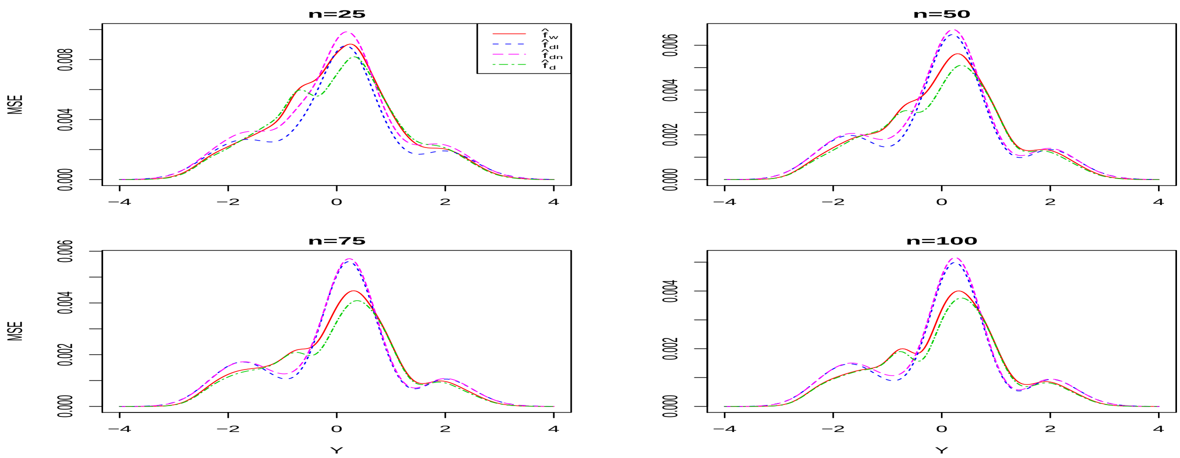

Figure 5.10: MSE for four KDEs under SRSWOR from the standard normal distribution - model III with $\sigma = 1.00$.

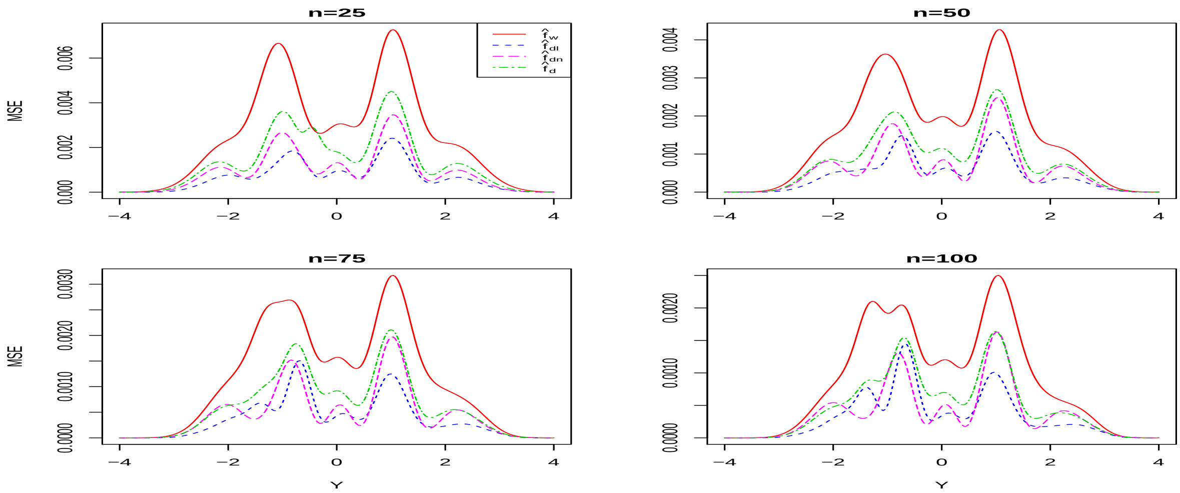

Figure 5.11: MSE for four KDEs under SRSWOR from the mixture normal distribution - model I with $\sigma = 0.20$.

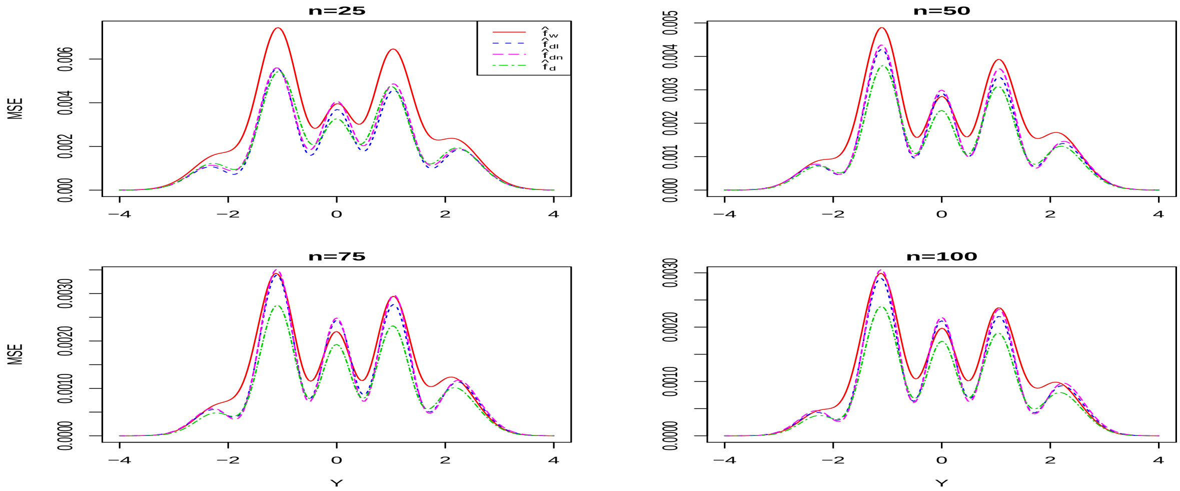

Figure 5.12: MSE for four KDEs under SRSWOR from the mixture normal distribution - model I with $\sigma = 0.60$.

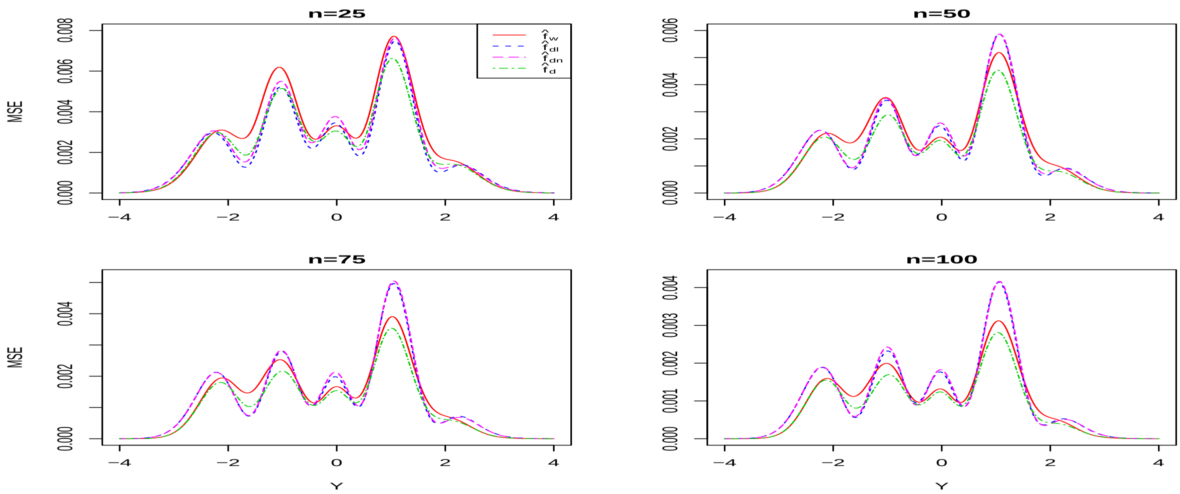

Figure 5.13: MSE for four KDEs under SRSWOR from the mixture normal distribution - model I with $\sigma = 1.00$.

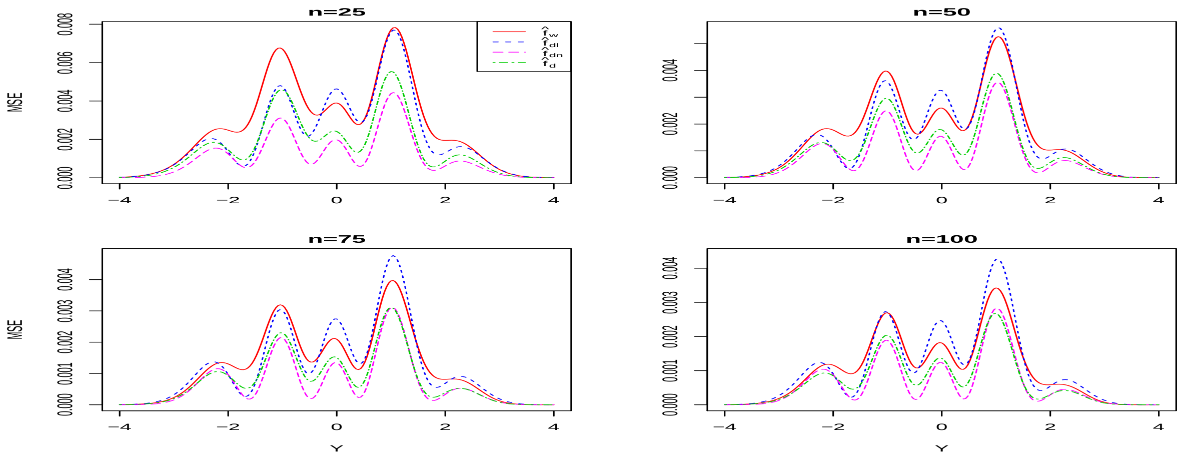

Figure 5.14: MSE for four KDEs under SRSWOR from the mixture normal distribution - model II with $\sigma = 0.20$.

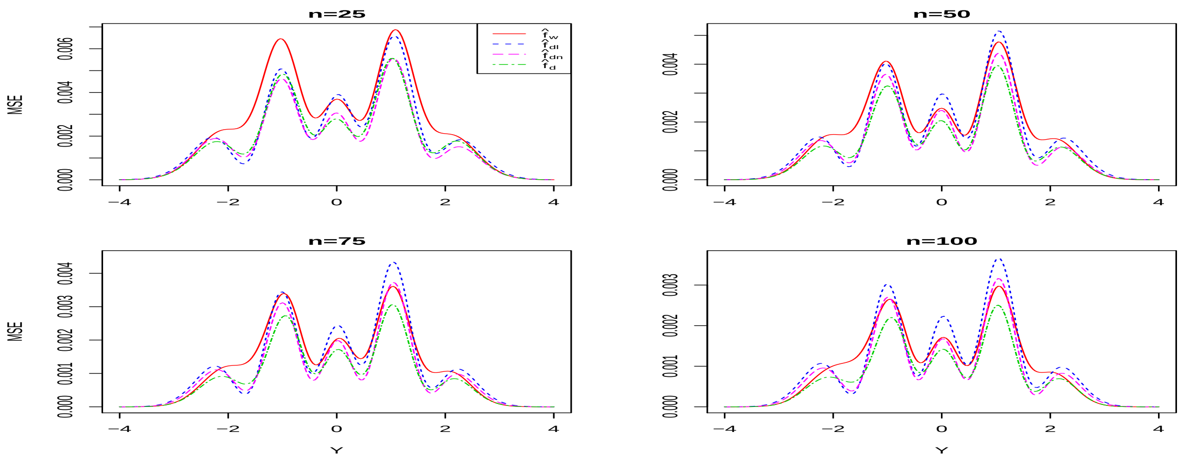

Figure 5.15: MSE for four KDEs under SRSWOR from the mixture normal distribution - model II with $\sigma = 0.60$.

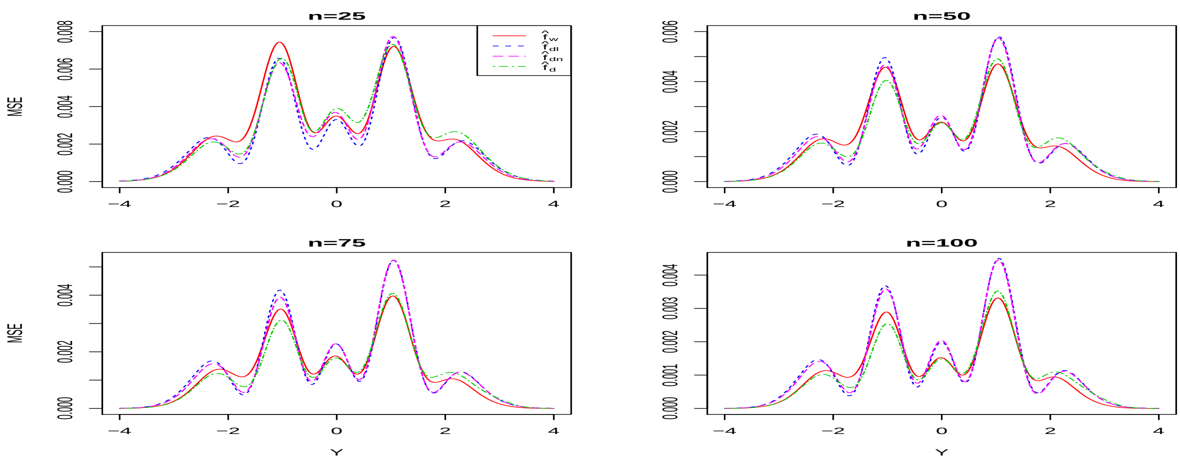

Figure 5.16: MSE for four KDEs under SRSWOR from the mixture normal distribution - model II with $\sigma = 1.00$.

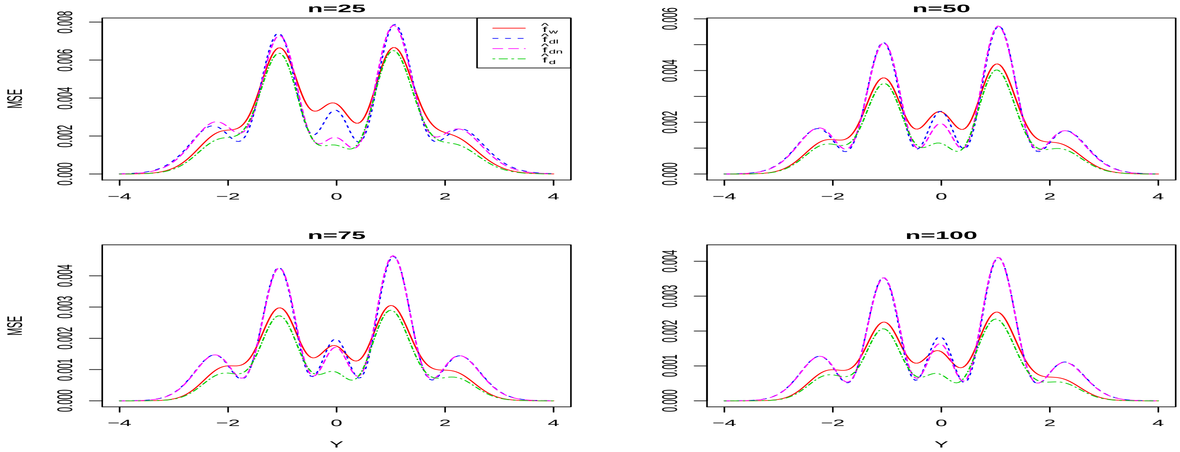

Figure 5.17: MSE for four KDEs under SRSWOR from the mixture normal distribution - model III with σ = 0.20.

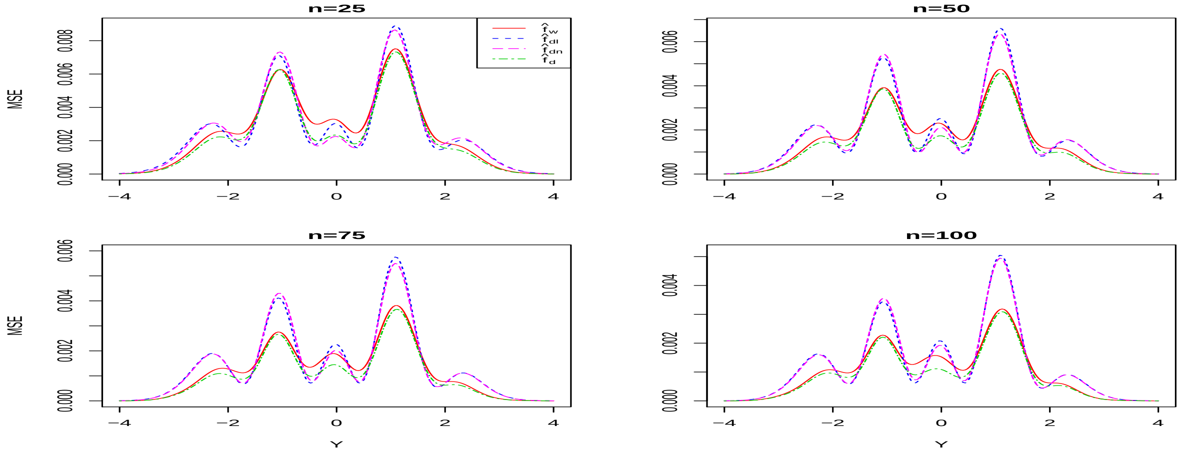

Figure 5.18: MSE for four KDEs under SRSWOR from the mixture normal distribution - model III with σ = 0.60.

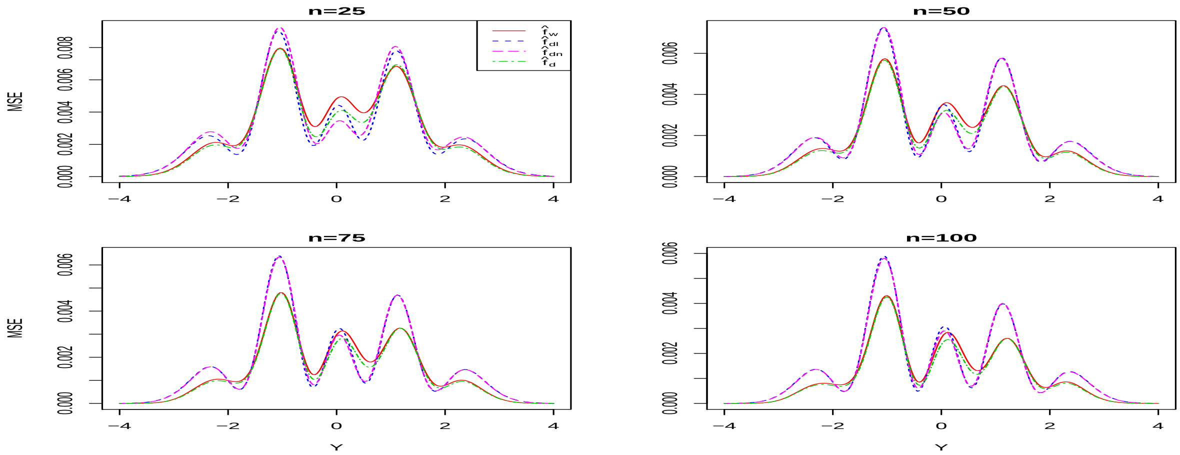

Figure 5.19: MSE for four KDEs under SRSWOR from the mixture normal distribution - model III with σ = 1.00.

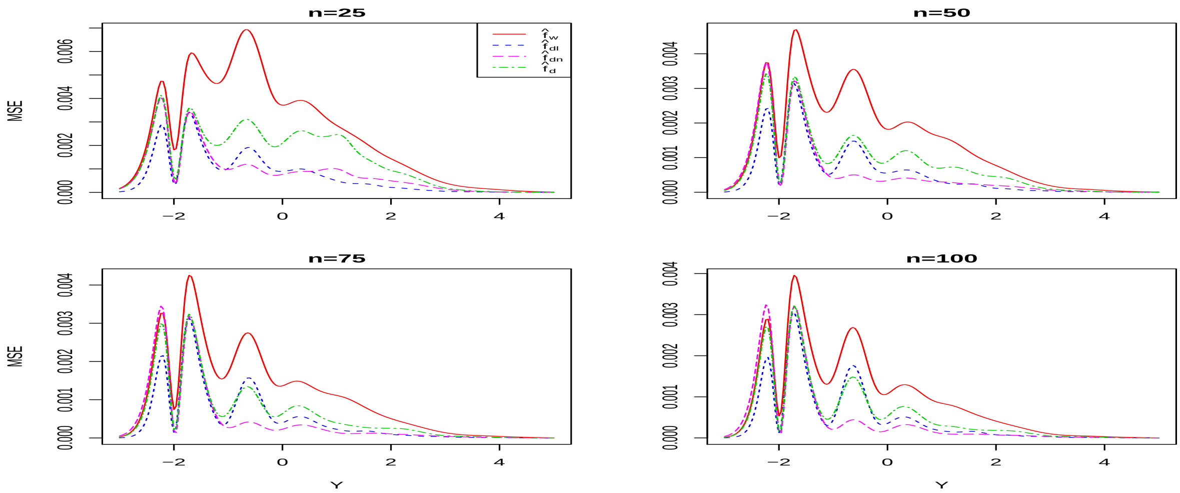

Figure 5.20: MSE for four KDEs under SRSWOR from the skew normal distribution - model I with $\sigma = 0.20$.

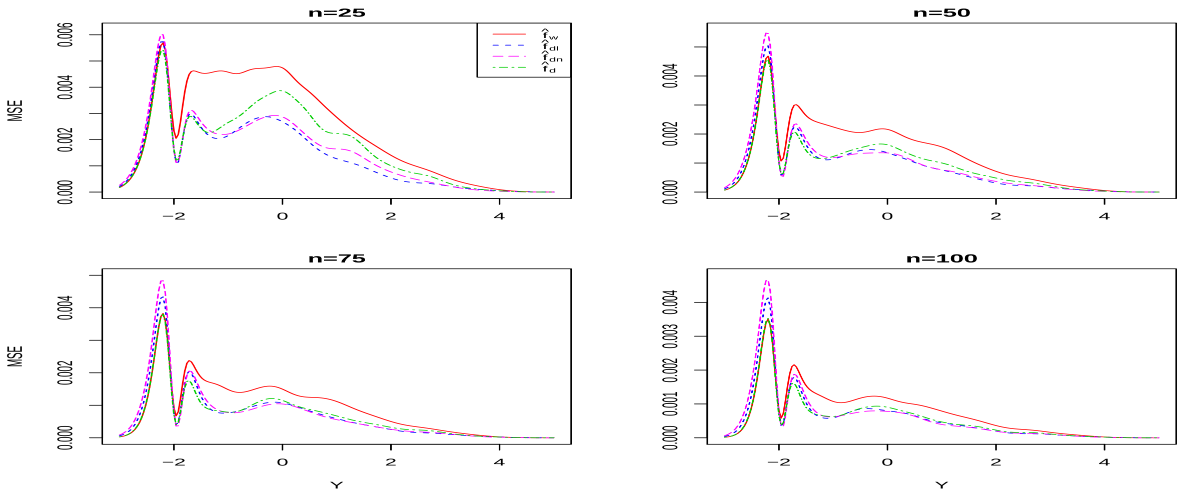

Figure 5.21: MSE for four KDEs under SRSWOR from the skew normal distribution - model I with $\sigma = 0.60$.

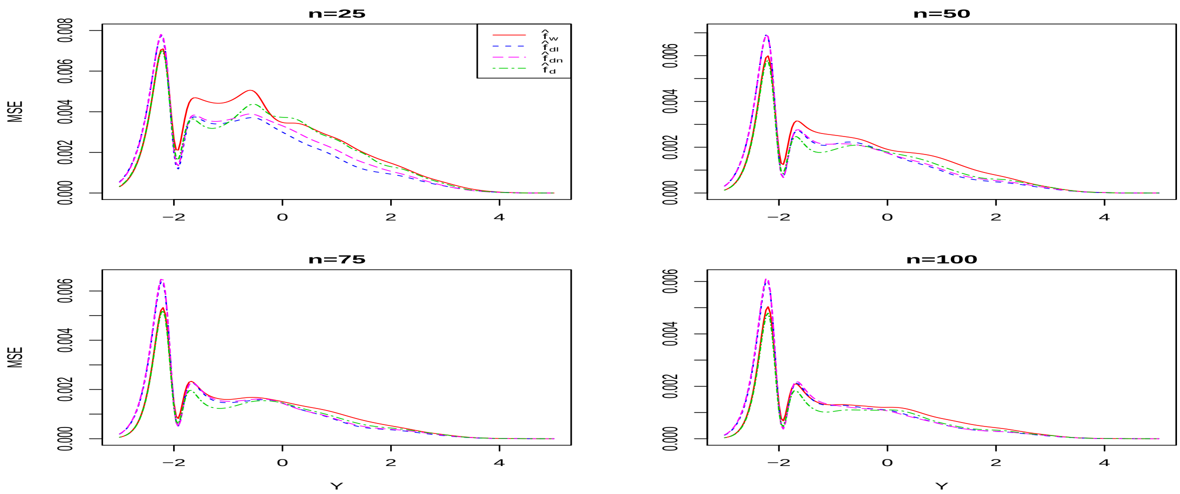

Figure 5.22: MSE for four KDEs under SRSWOR from the skew normal distribution - model I with $\sigma = 1.00$.

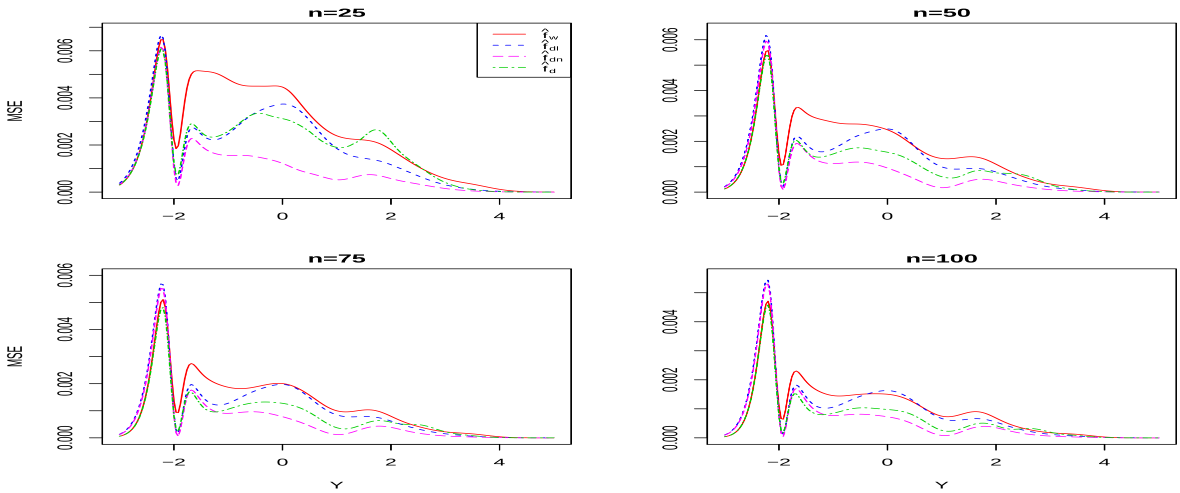

Figure 5.23: MSE for four KDEs under SRSWOR from the skew normal distribution - model II with $\sigma = 0.20$.

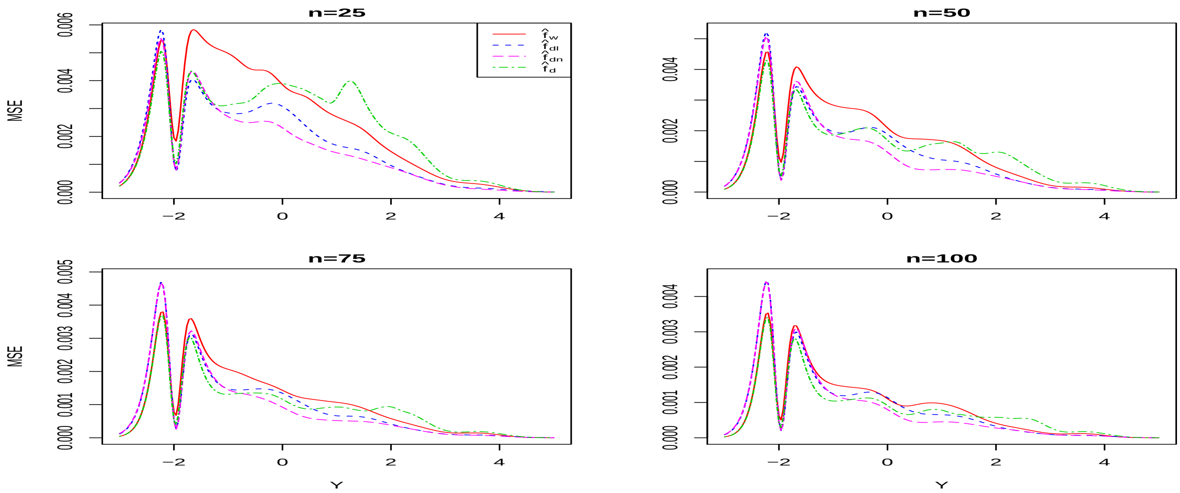

Figure 5.24: MSE for four KDEs under SRSWOR from the skew normal distribution - model II with $\sigma = 0.60$.

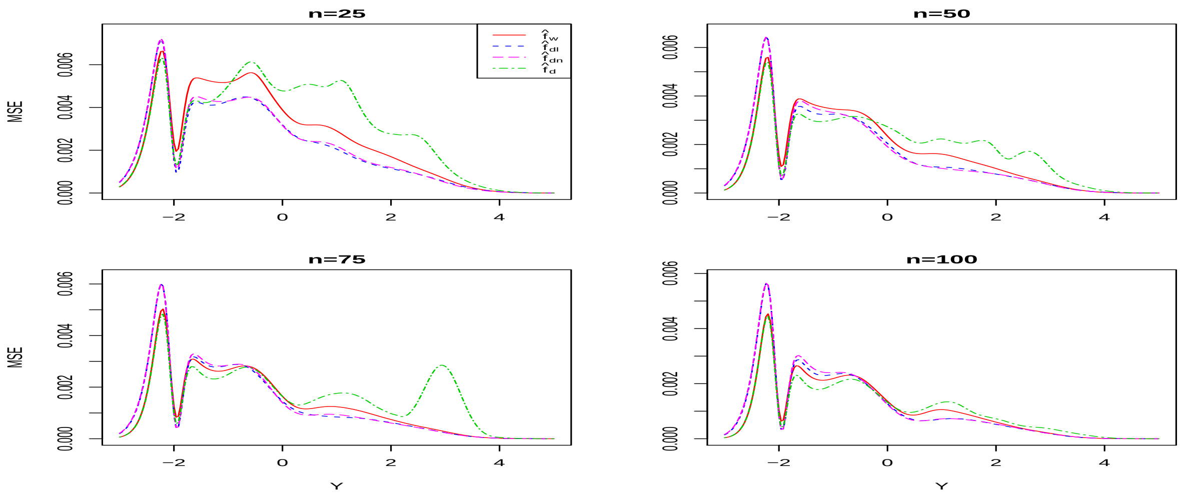

Figure 5.25: MSE for four KDEs under SRSWOR from the skew normal distribution - model II with $\sigma = 1.00$.

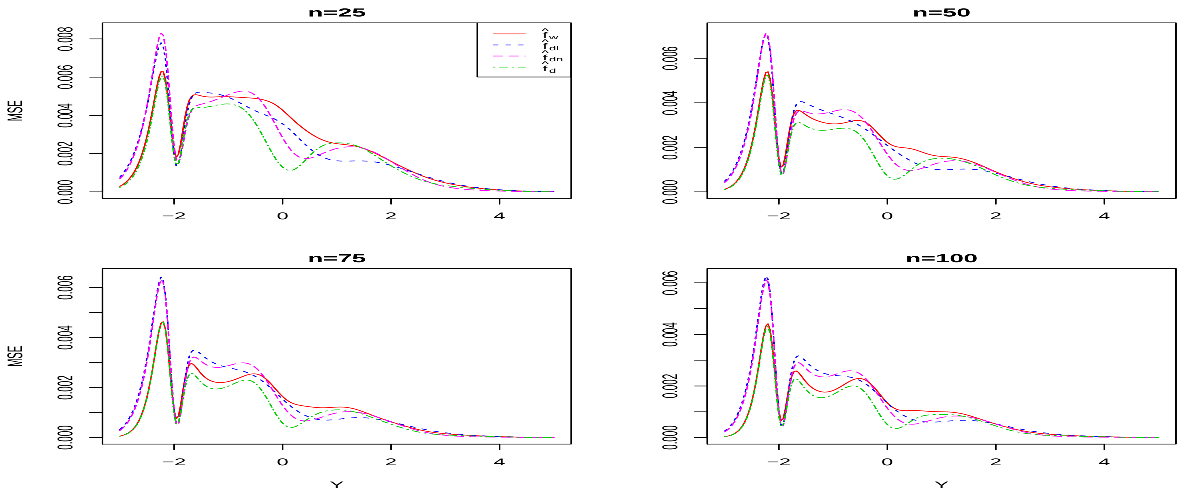

Figure 5.26: MSE for four KDEs under SRSWOR from the skew normal distribution - model III with $\sigma = 0.20$.

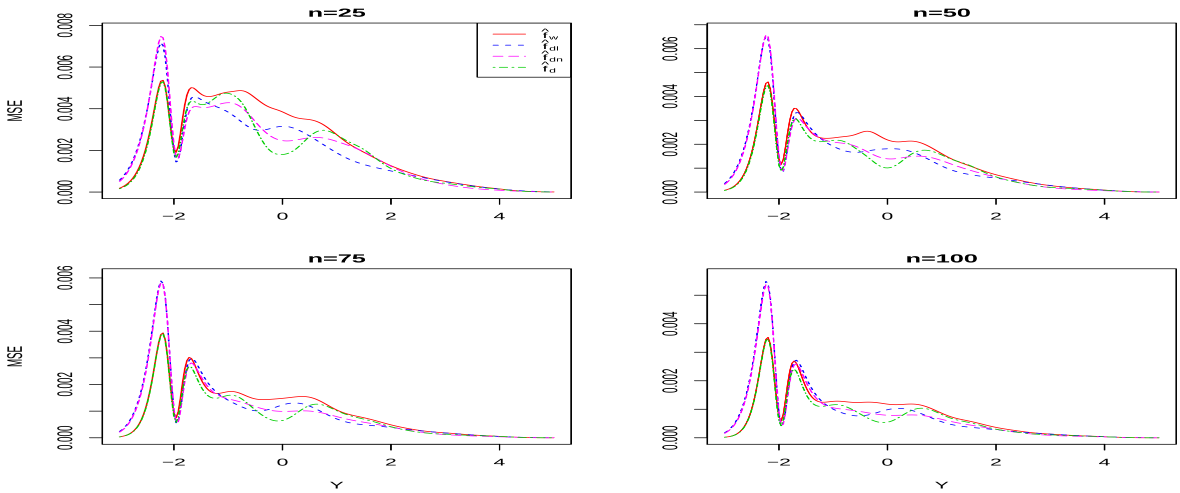

Figure 5.27: MSE for four KDEs under SRSWOR from the skew normal distribution - model III with $\sigma = 0.60$.

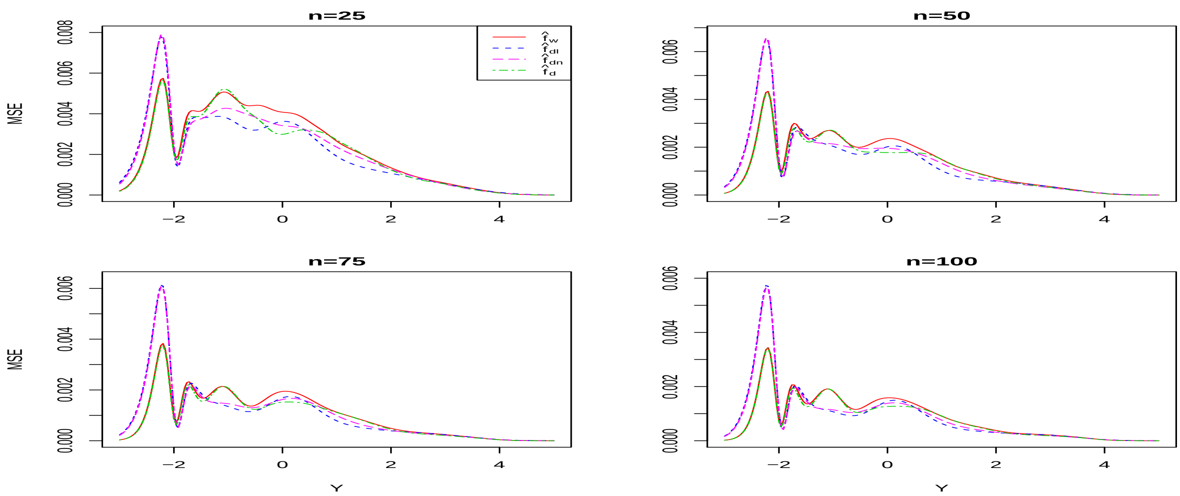

Figure 5.28: MSE for four KDEs under SRSWOR from the skew normal distribution - model III with $\sigma = 1.00$.

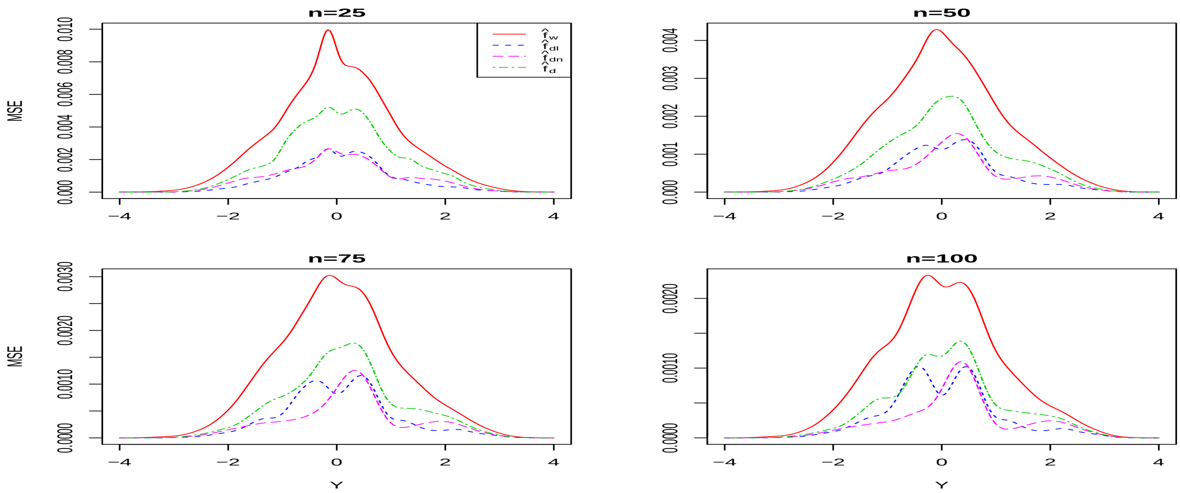

Figure 5.29: MSE for four KDEs under Poisson sampling from the standard normal distribution - model I with $\sigma = 0.20$.

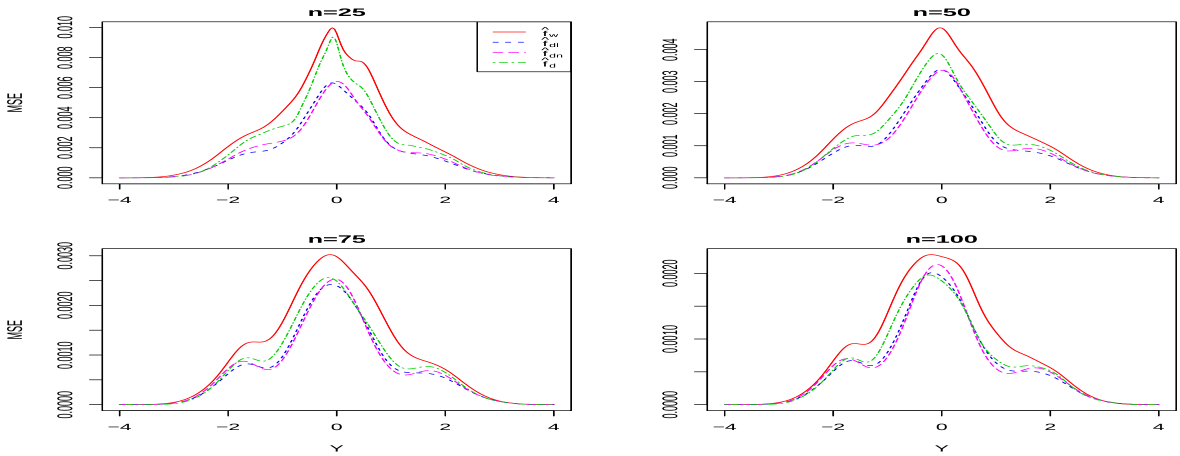

Figure 5.30: MSE for four KDEs under Poisson sampling from the standard normal distribution - model I with $\sigma = 0.60$.

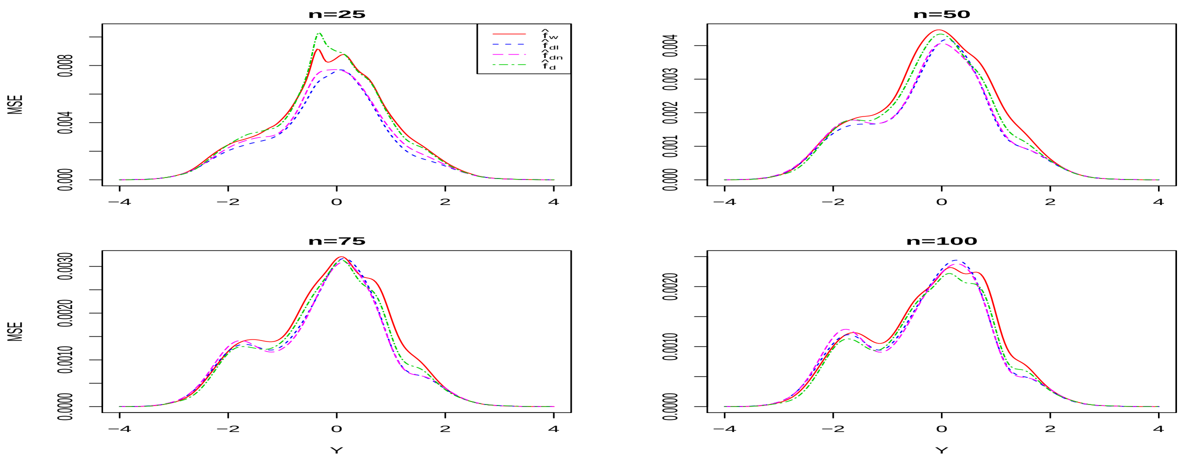

Figure 5.31: MSE for four KDEs under Poisson sampling from the standard normal distribution - model I with $\sigma = 1.00$.

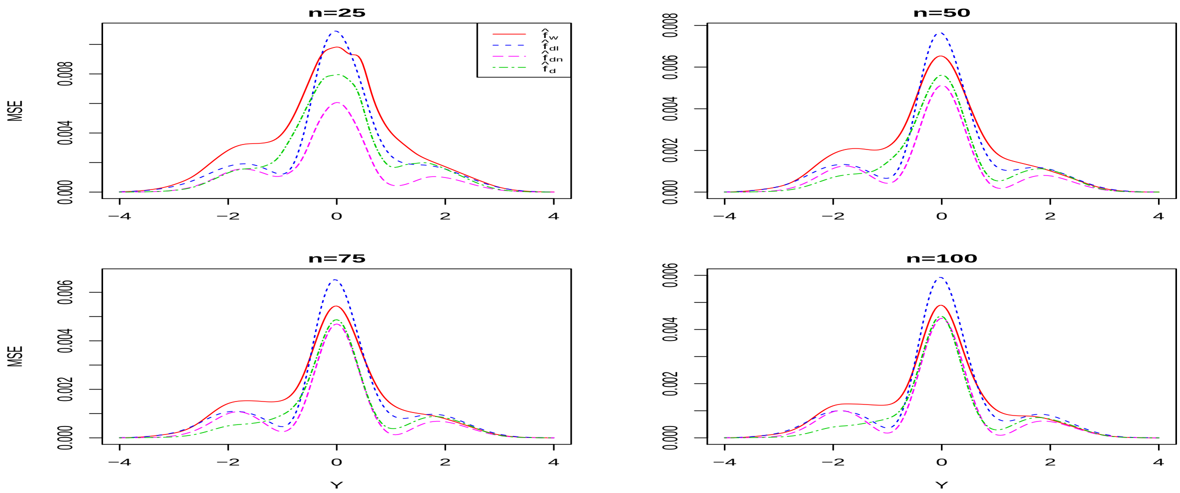

Figure 5.32: MSE for four KDEs under Poisson sampling from the standard normal distribution - model II with $\sigma = 0.20$.

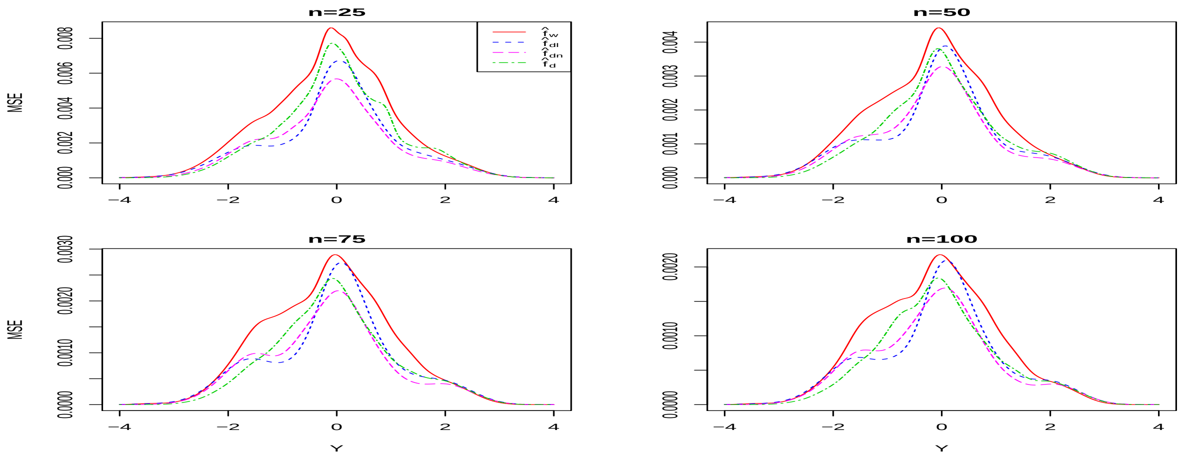

Figure 5.33: MSE for four KDEs under Poisson sampling from the standard normal distribution - model II with $\sigma = 0.60$.

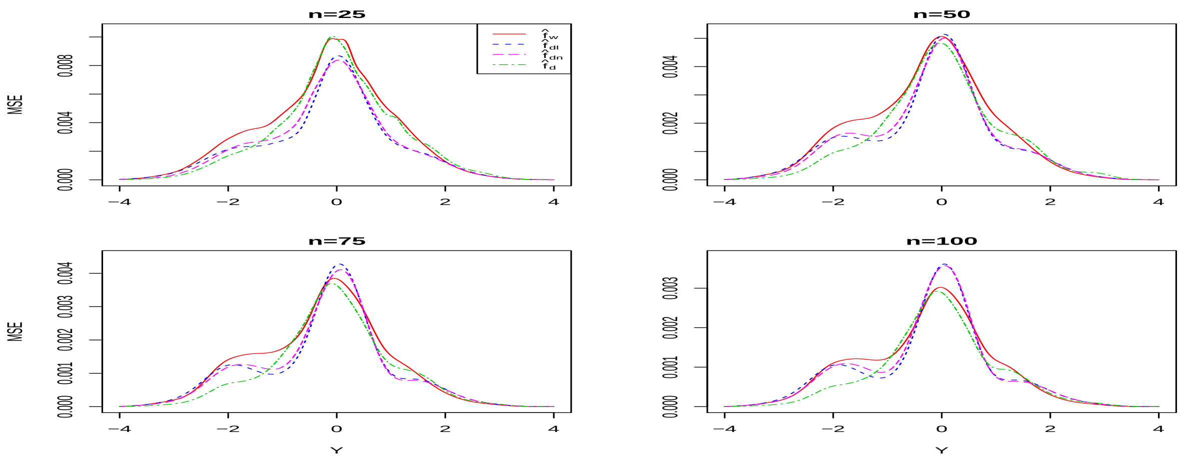

Figure 5.34: MSE for four KDEs under Poisson sampling from the standard normal distribution - model II with $\sigma = 1.00$.

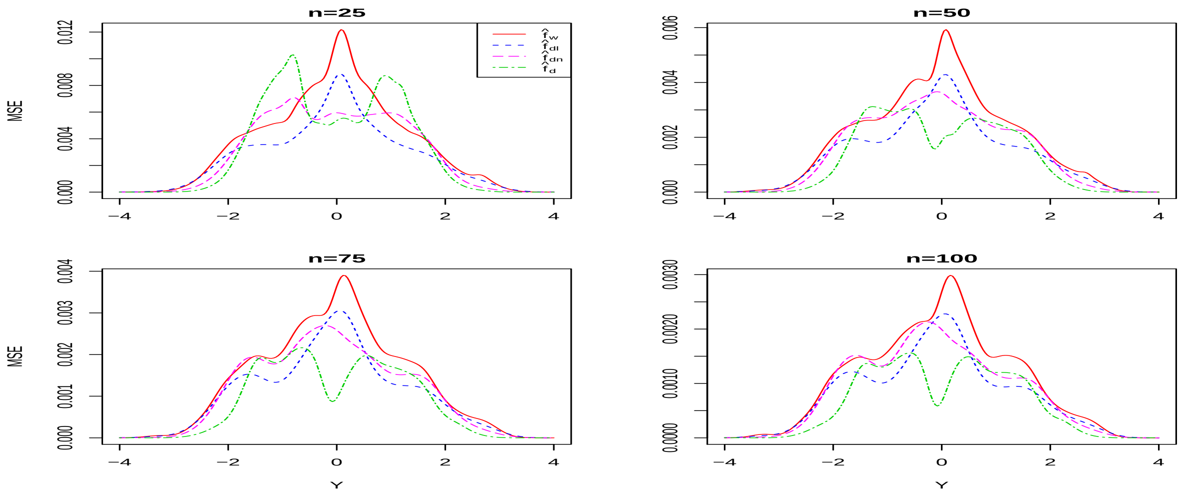

Figure 5.35: MSE for four KDEs under Poisson sampling from the standard normal distribution - model III with $\sigma = 0.20$.

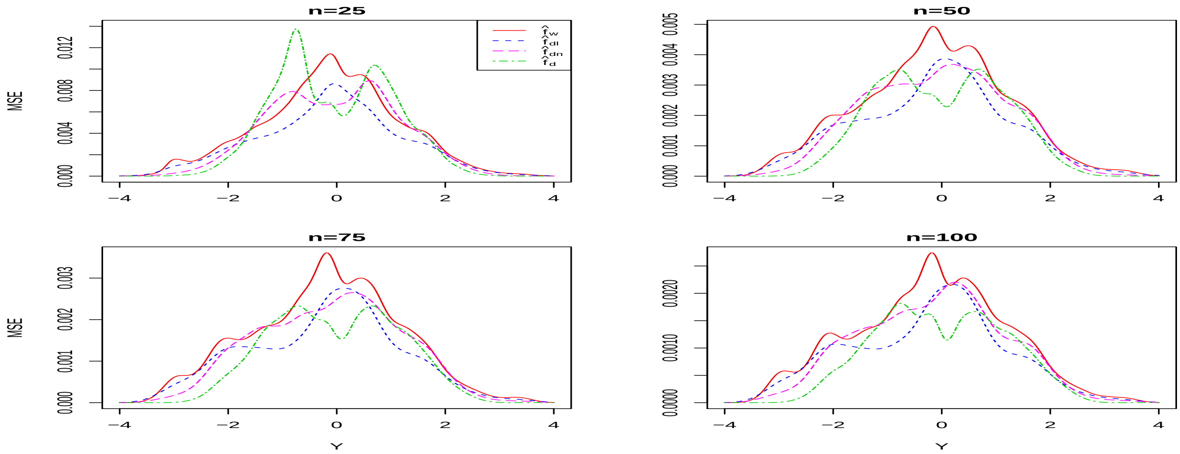

Figure 5.36: MSE for four KDEs under Poisson sampling from the standard normal distribution - model III with $\sigma = 0.60$.

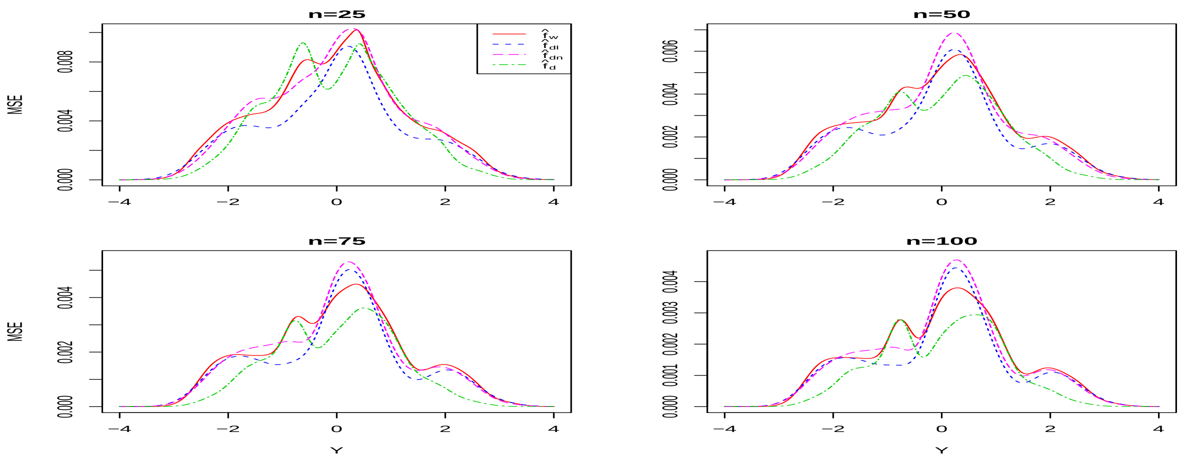

Figure 5.37: MSE for four KDEs under Poisson sampling from the standard normal distribution - model III with $\sigma = 1.00$.

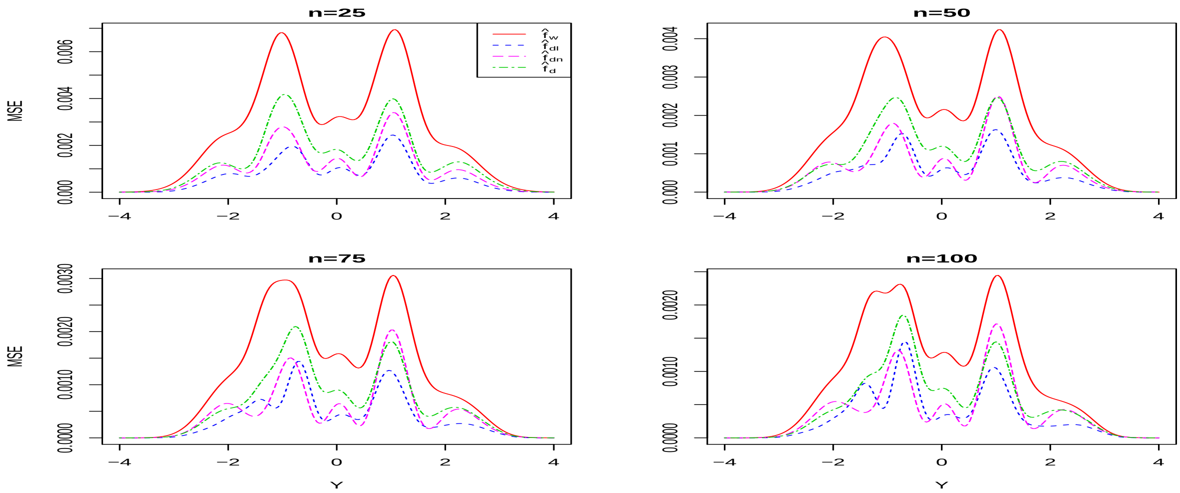

Figure 5.38: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model I with $\sigma = 0.20$.

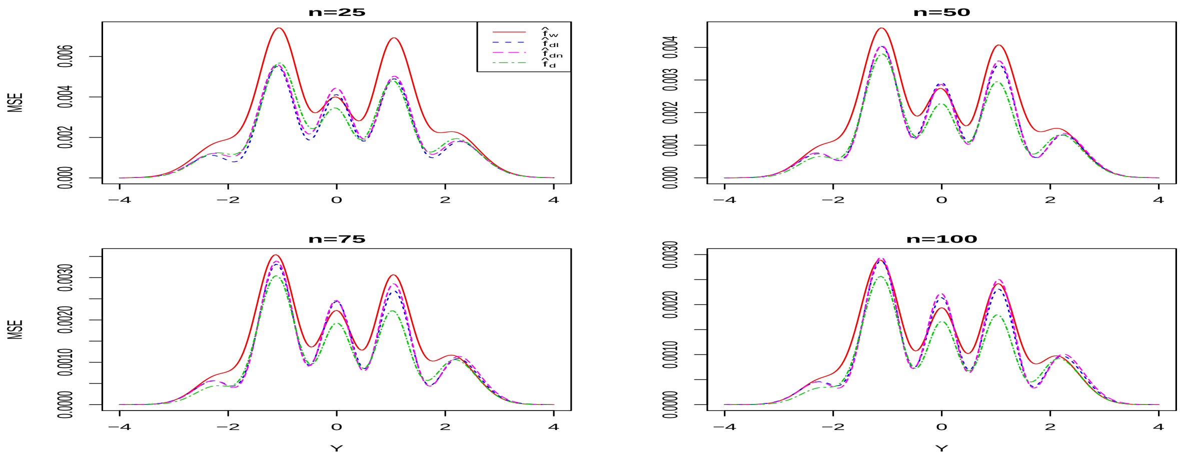

Figure 5.39: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model I with $\sigma = 0.60$.

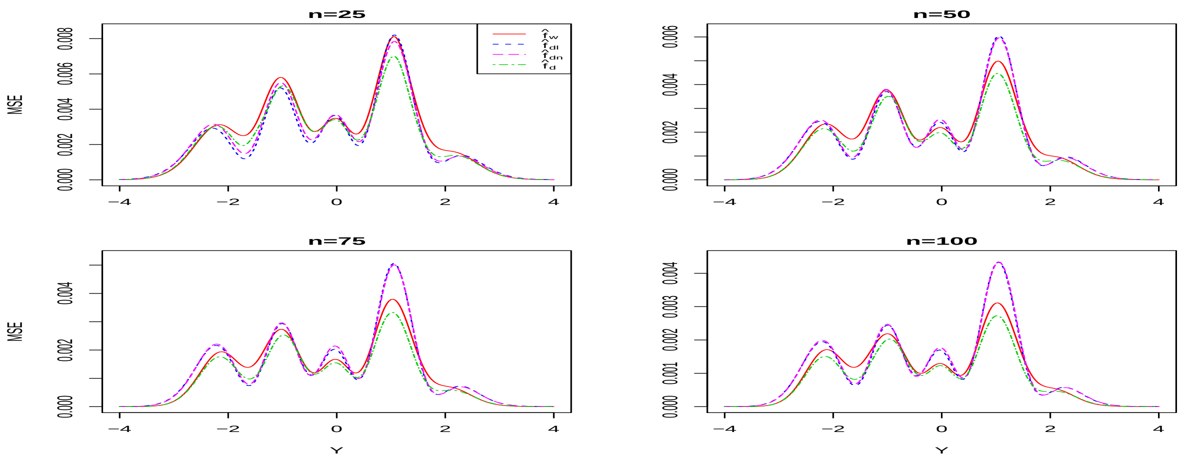

Figure 5.40: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model I with $\sigma = 1.00$.

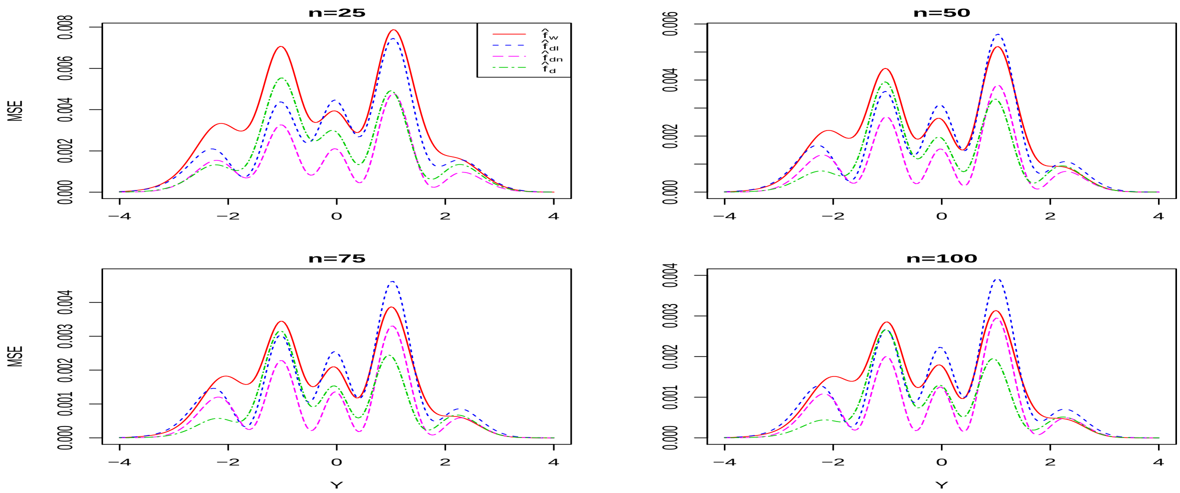

Figure 5.41: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model II with $\sigma = 0.20$.

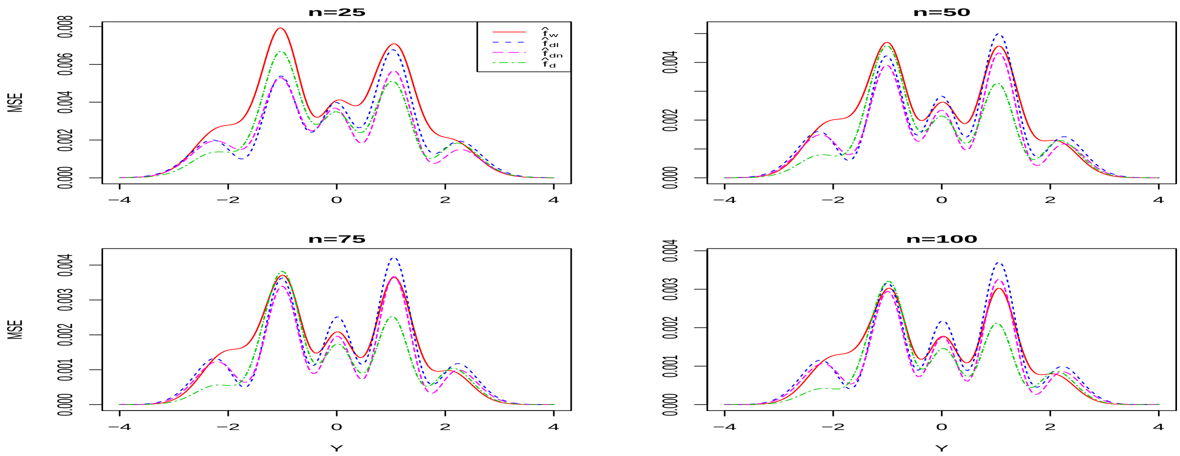

Figure 5.42: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model II with $\sigma = 0.60$.

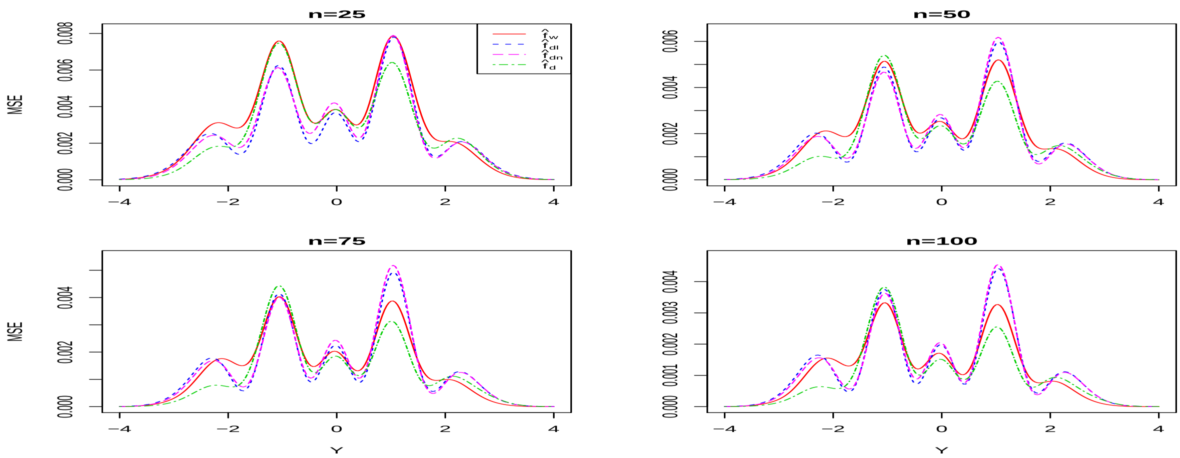

Figure 5.43: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model II with $\sigma = 1.00$.

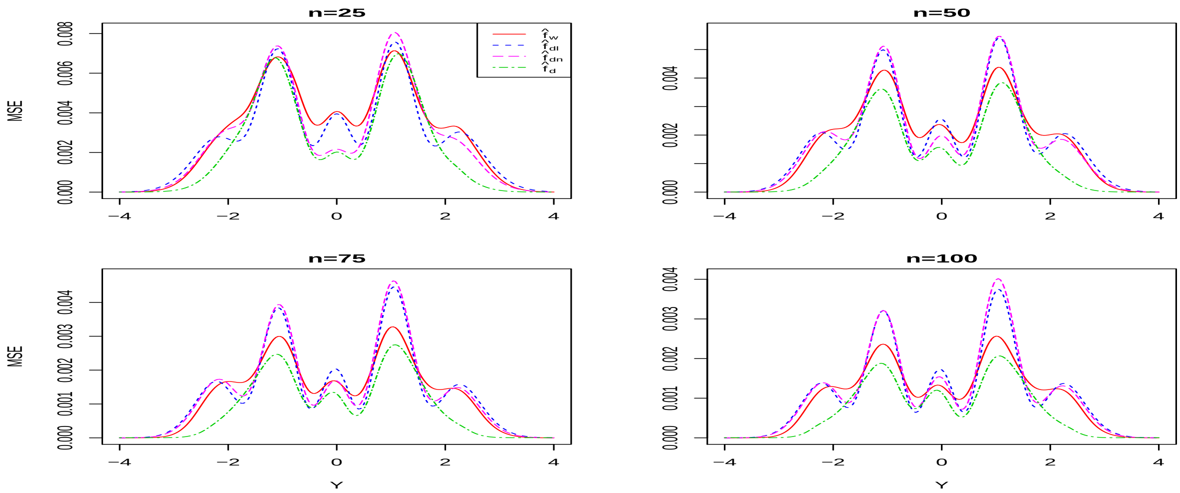

Figure 5.44: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model III with $\sigma = 0.20$.

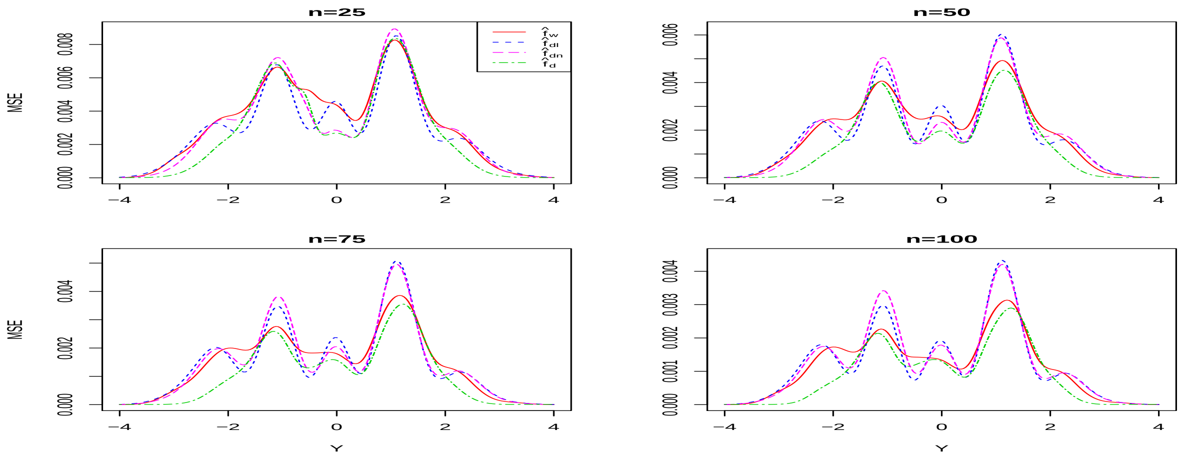

Figure 5.45: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model III with $\sigma = 0.60$.

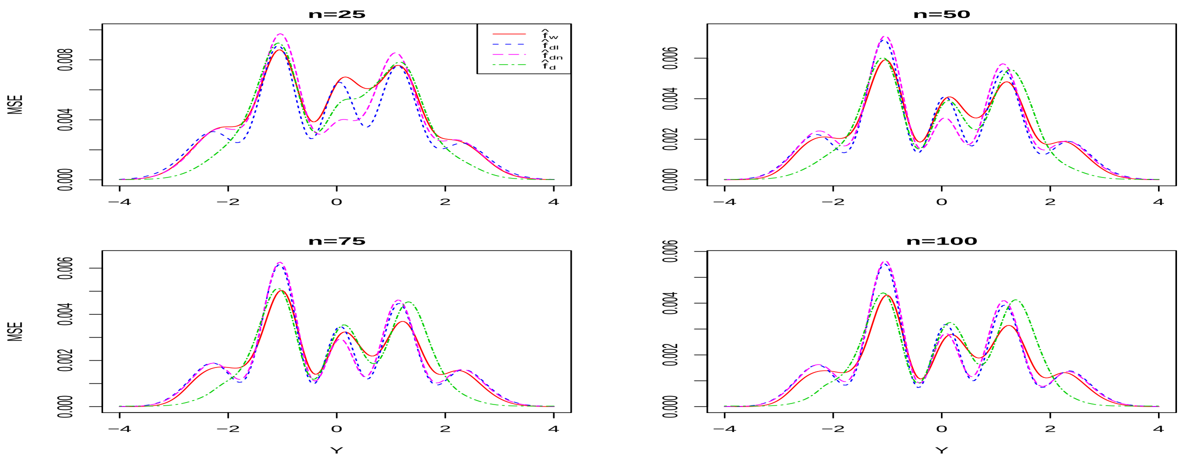

Figure 5.46: MSE for four KDEs under Poisson sampling from the mixture normal distribution - model III with $\sigma = 1.00$.

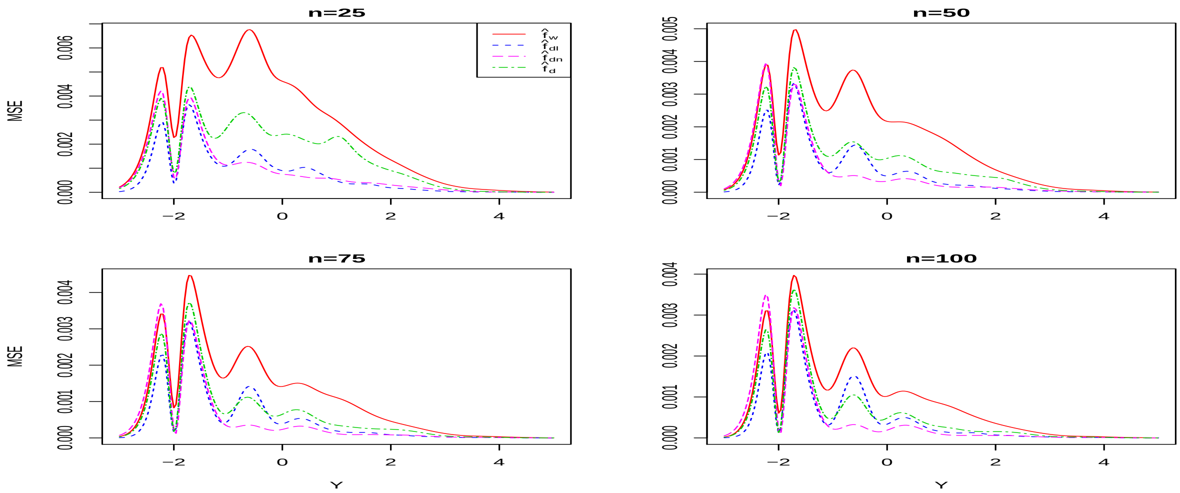

Figure 5.47: MSE for four KDEs under Poisson sampling from the skew normal distribution - model I with $\sigma = 0.20$.

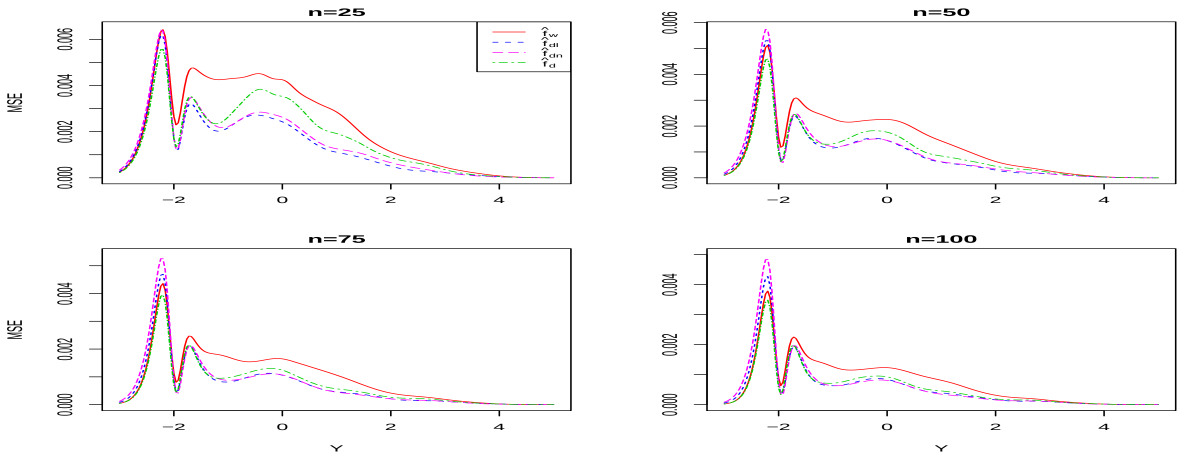

Figure 5.48: MSE for four KDEs under Poisson sampling from the skew normal distribution - model I with $\sigma = 0.60$.

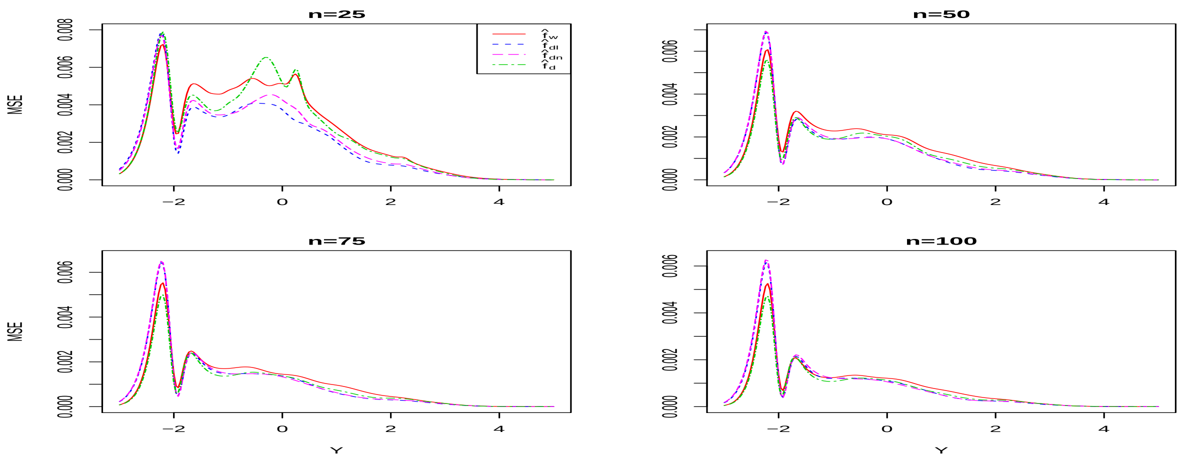

Figure 5.49: MSE for four KDEs under Poisson sampling from the skew normal distribution - model I with $\sigma = 1.00$.

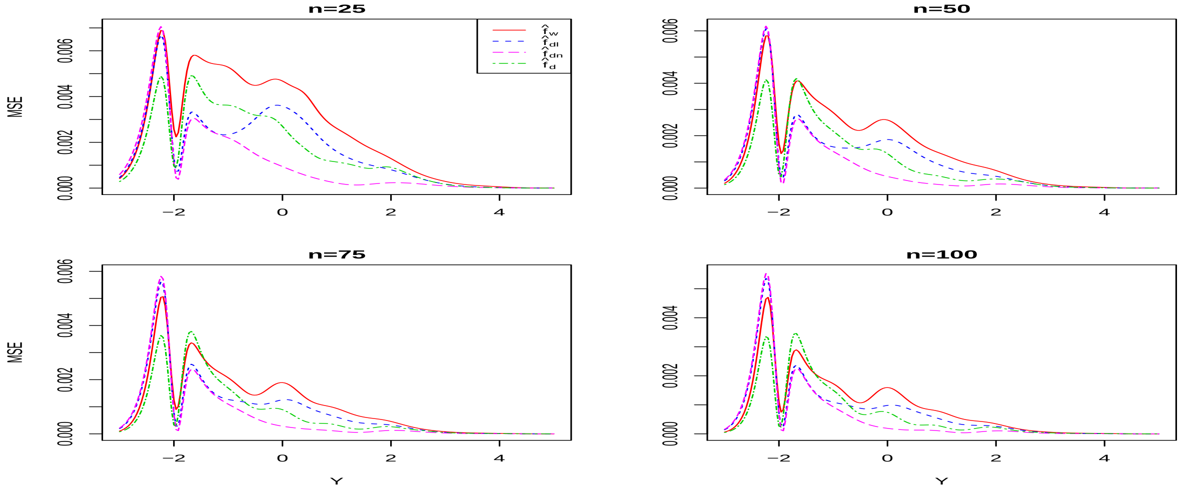

Figure 5.50: MSE for four KDEs under Poisson sampling from the skew normal distribution - model II with $\sigma = 0.20$.

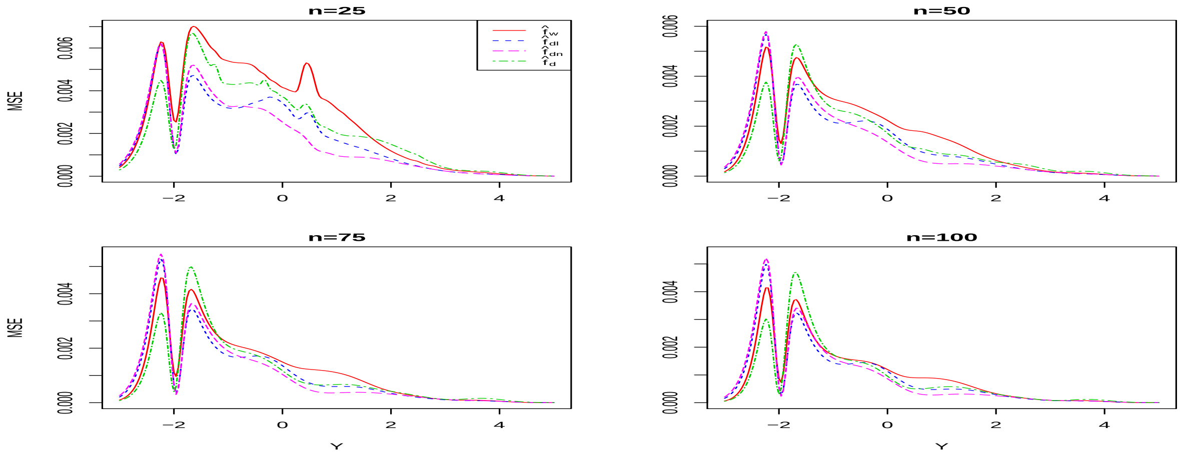

Figure 5.51: MSE for four KDEs under Poisson sampling from the skew normal distribution - model II with $\sigma = 0.60$.

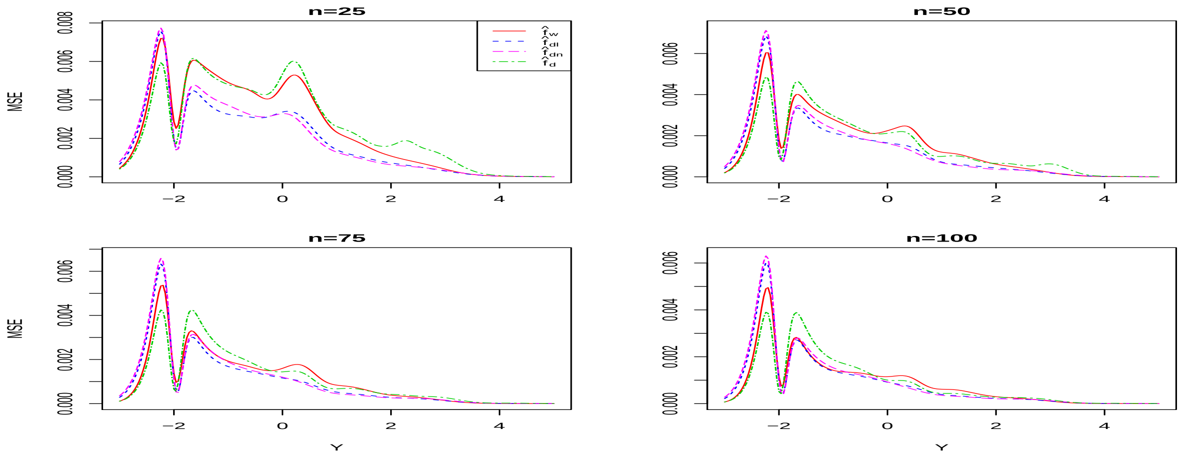

Figure 5.52: MSE for four KDEs under Poisson sampling from the skew normal distribution - model II with $\sigma = 1.00$.

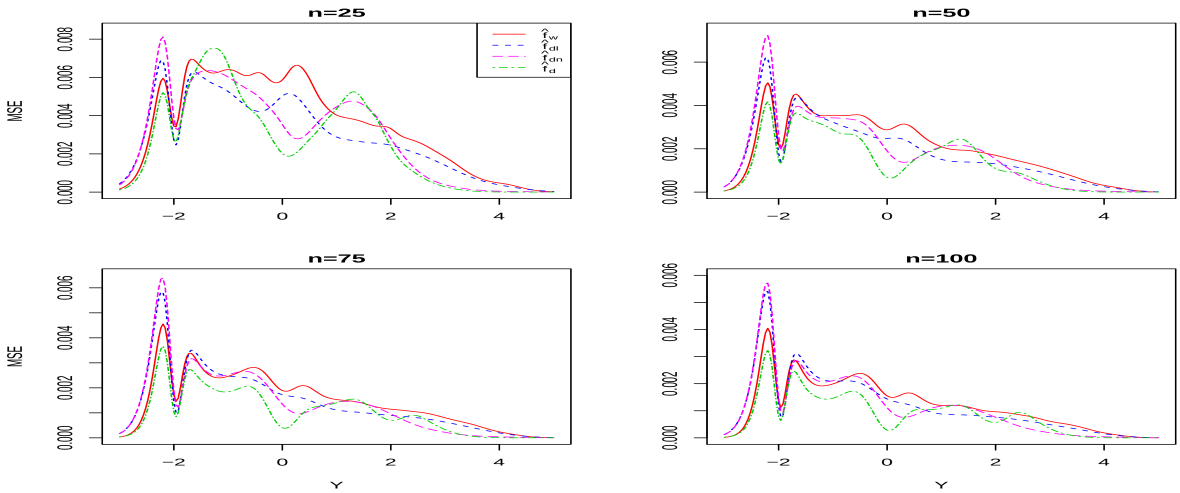

Figure 5.53: MSE for four KDEs under Poisson sampling from the skew normal distribution - model III with $\sigma = 0.20$.

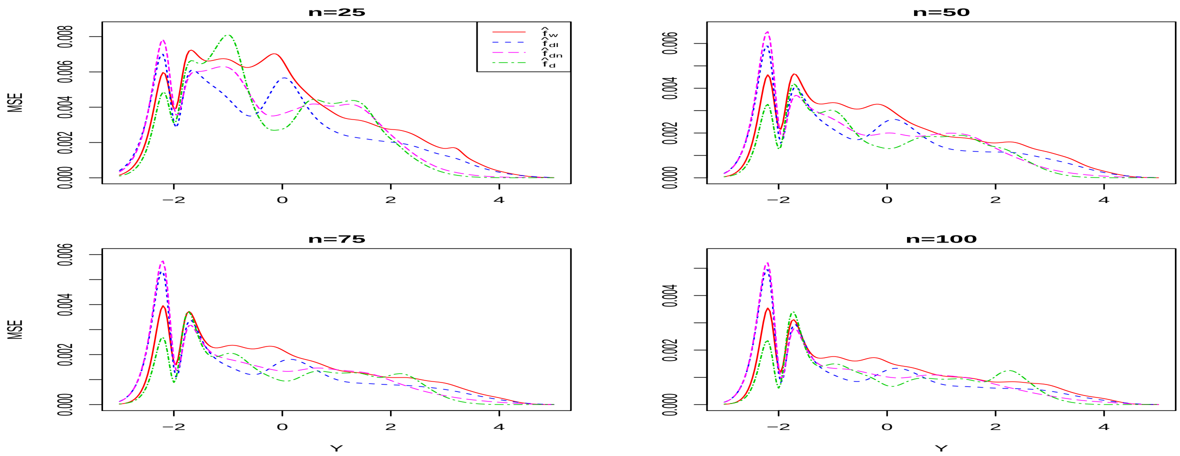

Figure 5.54: MSE for four KDEs under Poisson sampling from the skew normal distribution - model III with $\sigma = 0.60$.

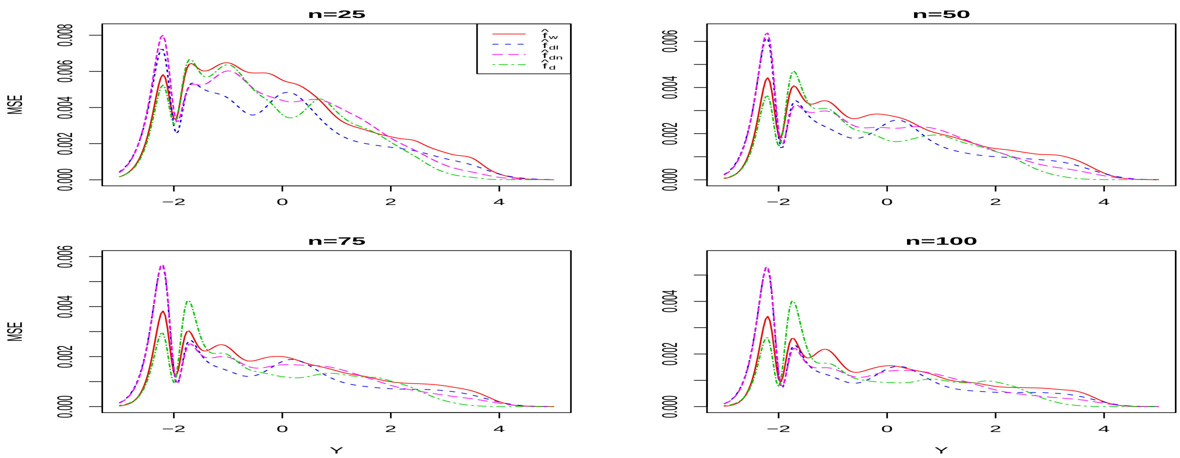

Figure 5.55: MSE for four KDEs under Poisson sampling from the skew normal distribution - model III with $\sigma = 1.00$.

Table 5.1: Monte Carlo MISE of four KDEs under SRSWOR from the standard normal distribution.

n	Est.	Model I			Model II			Model III		
		$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$
25	$\hat{f}_w$	0.01910	0.02159	0.02051	0.02131	0.02050	0.02028	0.01988	0.01926	0.02227
	$\hat{f}_{dl}$	0.00500	0.01262	0.01658	0.01732	0.01402	0.01577	0.01625	0.01554	0.01950
	$\hat{f}_{dn}$	0.00529	0.01302	0.01728	0.00903	0.01299	0.01693	0.01694	0.01716	0.02268
	$\hat{f}_d$	0.01128	0.01673	0.01974	0.01592	0.01786	0.02254	0.01528	0.01599	0.02117
50	$\hat{f}_w$	0.01023	0.01106	0.01115	0.01249	0.01040	0.01139	0.01040	0.01054	0.01348
	$\hat{f}_{dl}$	0.00296	0.00731	0.00950	0.01180	0.00761	0.01002	0.00914	0.00892	0.01313
	$\hat{f}_{dn}$	0.00276	0.00719	0.00963	0.00689	0.00700	0.01022	0.00953	0.00925	0.01392
	$\hat{f}_d$	0.00549	0.00832	0.01040	0.00927	0.00884	0.01194	0.00750	0.00852	0.01250
75	$\hat{f}_w$	0.00752	0.00776	0.00812	0.01011	0.00700	0.00815	0.00721	0.00752	0.01026
	$\hat{f}_{dl}$	0.00241	0.00560	0.00740	0.01010	0.00534	0.00778	0.00683	0.00680	0.01072
	$\hat{f}_{dn}$	0.00210	0.00562	0.00723	0.00612	0.00479	0.00796	0.00723	0.00711	0.01105
	$\hat{f}_d$	0.00389	0.00607	0.00752	0.00722	0.00572	0.00835	0.00506	0.00613	0.00952
100	$\hat{f}_w$	0.00563	0.00611	0.00625	0.00814	0.00554	0.00654	0.00570	0.00592	0.00880
	$\hat{f}_{dl}$	0.00207	0.00431	0.00587	0.00845	0.00449	0.00656	0.00537	0.00566	0.00920
	$\hat{f}_{dn}$	0.00169	0.00436	0.00585	0.00559	0.00395	0.00654	0.00557	0.00576	0.00957
	$\hat{f}_d$	0.00292	0.00468	0.00583	0.00605	0.00456	0.00643	0.00404	0.00474	0.00830

Table 5.2: Monte Carlo MISE of four KDEs under SRSWOR from the mixture normal distribution.

n	Est.	Model I			Model II			Model III		
		$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$
25	$\hat{f}_w$	0.02006	0.02000	0.01987	0.02097	0.01925	0.02035	0.01964	0.01972	0.02167
	$\hat{f}_{dl}$	0.00531	0.01351	0.01706	0.01712	0.01575	0.01806	0.01979	0.02019	0.02149
	$\hat{f}_{dn}$	0.00742	0.01442	0.01813	0.00871	0.01362	0.01864	0.01931	0.02031	0.02235
	$\hat{f}_d$	0.01103	0.01414	0.01707	0.01266	0.01414	0.02017	0.01615	0.01759	0.02037
50	$\hat{f}_w$	0.01187	0.01221	0.01244	0.01272	0.01216	0.01275	0.01138	0.01228	0.01393
	$\hat{f}_{dl}$	0.00357	0.00962	0.01199	0.01193	0.01143	0.01292	0.01284	0.01411	0.01527
	$\hat{f}_{dn}$	0.00482	0.01000	0.01225	0.00645	0.00965	0.01280	0.01278	0.01399	0.01531
	$\hat{f}_d$	0.00646	0.00892	0.01056	0.00822	0.00904	0.01235	0.00932	0.01085	0.01316
75	$\hat{f}_w$	0.00906	0.00885	0.00941	0.00966	0.00949	0.00987	0.00865	0.00918	0.01096
	$\hat{f}_{dl}$	0.00296	0.00775	0.00991	0.00973	0.00940	0.01098	0.01037	0.01138	0.01271
	$\hat{f}_{dn}$	0.00380	0.00788	0.01008	0.00551	0.00797	0.01086	0.01036	0.01130	0.01266
	$\hat{f}_d$	0.00506	0.00664	0.00827	0.00639	0.00721	0.00962	0.00700	0.00807	0.01036
100	$\hat{f}_w$	0.00704	0.00746	0.00745	0.00800	0.00756	0.00810	0.00673	0.00764	0.00909
	$\hat{f}_{dl}$	0.00258	0.00643	0.00827	0.00857	0.00797	0.00933	0.00864	0.00972	0.01106
	$\hat{f}_{dn}$	0.00308	0.00662	0.00838	0.00487	0.00667	0.00929	0.00863	0.00974	0.01100
	$\hat{f}_d$	0.00395	0.00565	0.00656	0.00546	0.00577	0.00794	0.00547	0.00676	0.00860

Table 5.3: Monte Carlo MISE of four KDEs under SRSWOR from the skew normal distribution.

n	Est.	Model I			Model II			Model III		
		$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$
25	$\hat{f}_w$	0.01930	0.01853	0.01826	0.01908	0.01853	0.01981	0.01882	0.01773	0.01783
	$\hat{f}_{dl}$	0.00558	0.01083	0.01484	0.01380	0.01323	0.01636	0.01794	0.01570	0.01619
	$\hat{f}_{dn}$	0.00657	0.01178	0.01585	0.00780	0.01193	0.01670	0.01804	0.01613	0.01718
	$\hat{f}_d$	0.01139	0.01363	0.01657	0.01429	0.01770	0.02414	0.01493	0.01495	0.01652
50	$\hat{f}_w$	0.01105	0.01002	0.01086	0.01174	0.01134	0.01273	0.01210	0.01061	0.01020
	$\hat{f}_{dl}$	0.00417	0.00680	0.00980	0.00978	0.00936	0.01174	0.01270	0.01017	0.01015
	$\hat{f}_{dn}$	0.00427	0.00717	0.00994	0.00618	0.00827	0.01175	0.01249	0.01005	0.01041
	$\hat{f}_d$	0.00620	0.00716	0.00919	0.00795	0.01017	0.01435	0.00977	0.00882	0.00950
75	$\hat{f}_w$	0.00841	0.00718	0.00794	0.00914	0.00831	0.00990	0.00912	0.00765	0.00786
	$\hat{f}_{dl}$	0.00376	0.00531	0.00786	0.00806	0.00729	0.00987	0.01022	0.00778	0.00805
	$\hat{f}_{dn}$	0.00360	0.00556	0.00803	0.00530	0.00667	0.00998	0.01007	0.00754	0.00820
	$\hat{f}_d$	0.00475	0.00519	0.00705	0.00617	0.00721	0.01231	0.00757	0.00643	0.00726
100	$\hat{f}_w$	0.00728	0.00574	0.00664	0.00743	0.006800	0.00816	0.00785	0.00620	0.0066
	$\hat{f}_{dl}$	0.00357	0.00447	0.00667	0.00700	0.00647	0.00849	0.00915	0.00646	0.00691
	$\hat{f}_{dn}$	0.00335	0.00474	0.00674	0.00487	0.00591	0.00864	0.00887	0.00629	0.00695
	$\hat{f}_d$	0.00436	0.00427	0.00575	0.00515	0.00607	0.00823	0.00652	0.00520	0.00617

Table 5.4: Monte Carlo MISE of four KDEs under Poisson sampling from the standard normal distribution.

n	Est.	Model I			Model II			Model III		
		$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$
25	$\hat{f}_w$	0.02161	0.02200	0.02187	0.02372	0.02020	0.02349	0.02818	0.03014	0.02903
	$\hat{f}_{dl}$	0.00551	0.01324	0.01759	0.01837	0.01354	0.01766	0.02077	0.02262	0.02359
	$\hat{f}_{dn}$	0.00610	0.01340	0.01868	0.01016	0.01258	0.01809	0.02307	0.02601	0.02891
	$\hat{f}_d$	0.01270	0.01764	0.02202	0.01611	0.01588	0.02121	0.02523	0.02865	0.02594
50	$\hat{f}_w$	0.01077	0.01146	0.01159	0.01374	0.01047	0.01247	0.01463	0.01483	0.01732
	$\hat{f}_{dl}$	0.00302	0.00735	0.00977	0.01206	0.00767	0.01041	0.01096	0.01137	0.01490
	$\hat{f}_{dn}$	0.00307	0.00729	0.00991	0.00760	0.00717	0.01052	0.01193	0.01248	0.01727
	$\hat{f}_d$	0.00582	0.00863	0.01063	0.00922	0.00792	0.01069	0.00970	0.01085	0.01328
75	$\hat{f}_w$	0.00776	0.00786	0.00824	0.01058	0.00736	0.00911	0.01041	0.01050	0.01308
	$\hat{f}_{dl}$	0.00241	0.00551	0.00738	0.00993	0.00559	0.00821	0.00803	0.00806	0.01155
	$\hat{f}_{dn}$	0.00215	0.00555	0.00735	0.00663	0.00507	0.00818	0.00859	0.00879	0.01295
	$\hat{f}_d$	0.00392	0.00609	0.00755	0.00718	0.00541	0.00773	0.00657	0.00734	0.00954
100	$\hat{f}_w$	0.00601	0.00613	0.00631	0.00878	0.00560	0.00697	0.00796	0.00814	0.01080
	$\hat{f}_{dl}$	0.00205	0.00447	0.00592	0.00875	0.00428	0.00671	0.00611	0.00629	0.00970
	$\hat{f}_{dn}$	0.00171	0.00454	0.00591	0.00611	0.00399	0.00683	0.00664	0.00692	0.01080
	$\hat{f}_d$	0.00299	0.00476	0.00577	0.00615	0.00415	0.00596	0.00489	0.00558	0.00784

Table 5.5: Monte Carlo MISE of four KDEs under Poisson sampling from the mixture normal distribution.

n	Est.	Model I			Model II			Model III		
		$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$
25	$\hat{f}_w$	0.02026	0.02044	0.01968	0.02197	0.02229	0.02240	0.02345	0.02576	0.02807
	$\hat{f}_{dl}$	0.00551	0.01406	0.01755	0.01671	0.01684	0.01870	0.02169	0.02317	0.02513
	$\hat{f}_{dn}$	0.00771	0.01494	0.01802	0.00918	0.01508	0.01915	0.02159	0.02416	0.02659
	$\hat{f}_d$	0.01076	0.01460	0.01759	0.01311	0.01603	0.01958	0.01741	0.02119	0.02472
50	$\hat{f}_w$	0.01237	0.01230	0.01245	0.01327	0.01337	0.01433	0.01405	0.01536	0.01709
	$\hat{f}_{dl}$	0.00366	0.00972	0.01237	0.01167	0.01173	0.01345	0.01409	0.01508	0.01662
	$\hat{f}_{dn}$	0.00490	0.00975	0.01252	0.00680	0.01014	0.01347	0.01406	0.01523	0.01705
	$\hat{f}_d$	0.00661	0.00896	0.01096	0.00825	0.00990	0.01241	0.00947	0.01146	0.01534
75	$\hat{f}_w$	0.00909	0.00925	0.00937	0.01014	0.01022	0.01077	0.01007	0.01130	0.01351
	$\hat{f}_{dl}$	0.00297	0.00773	0.01010	0.00951	0.00971	0.01086	0.01075	0.01162	0.01376
	$\hat{f}_{dn}$	0.00384	0.00786	0.01019	0.00579	0.00830	0.01108	0.01103	0.01180	0.01389
	$\hat{f}_d$	0.00500	0.00701	0.00818	0.00630	0.00773	0.00939	0.00678	0.00838	0.01271
100	$\hat{f}_w$	0.00716	0.00749	0.00765	0.00829	0.00825	0.00889	0.00794	0.00906	0.01126
	$\hat{f}_{dl}$	0.00262	0.00660	0.00854	0.00811	0.00823	0.00951	0.00882	0.00963	0.01190
	$\hat{f}_{dn}$	0.00319	0.00675	0.00867	0.00508	0.00714	0.00956	0.00911	0.00995	0.01219
	$\hat{f}_d$	0.00402	0.00578	0.00667	0.00515	0.00631	0.00771	0.00529	0.00691	0.01117

Table 5.6: Monte Carlo MISE of four KDEs under Poisson sampling from the skew normal distribution.

n	Est.	Model I			Model II			Model III		
		$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 0.6$	$\sigma = 1.0$
25	$\hat{f}_w$	0.02022	0.01785	0.02005	0.01964	0.02095	0.01975	0.02784	0.02713	0.02512
	$\hat{f}_{dl}$	0.00577	0.01071	0.01574	0.01303	0.01445	0.01476	0.02305	0.02158	0.02064
	$\hat{f}_{dn}$	0.00643	0.01159	0.01659	0.00823	0.01316	0.01504	0.02362	0.02302	0.02369
	$\hat{f}_d$	0.01173	0.01347	0.01965	0.01240	0.01733	0.02090	0.02074	0.02189	0.02130
50	$\hat{f}_w$	0.01164	0.01028	0.01061	0.01154	0.01201	0.01160	0.01591	0.01507	0.01409
	$\hat{f}_{dl}$	0.00429	0.00693	0.00966	0.00859	0.00960	0.00982	0.01430	0.01256	0.01231
	$\hat{f}_{dn}$	0.00433	0.00735	0.00983	0.00597	0.00910	0.01006	0.01427	0.01275	0.01335
	$\hat{f}_d$	0.00637	0.00749	0.00938	0.00751	0.01013	0.01140	0.01100	0.01070	0.01130
75	$\hat{f}_w$	0.00853	0.00765	0.00804	0.00854	0.00905	0.00867	0.01174	0.01110	0.01029
	$\hat{f}_{dl}$	0.00377	0.00545	0.00767	0.00687	0.00789	0.00793	0.01117	0.00950	0.00929
	$\hat{f}_{dn}$	0.00371	0.00580	0.00775	0.00506	0.00766	0.00828	0.01076	0.00983	0.00992
	$\hat{f}_d$	0.00483	0.00556	0.00701	0.00588	0.00798	0.00872	0.00803	0.00824	0.00849
100	$\hat{f}_w$	0.00691	0.00604	0.00657	0.00711	0.00742	0.00690	0.00962	0.00862	0.00835
	$\hat{f}_{dl}$	0.00351	0.00451	0.00663	0.00596	0.00689	0.00682	0.00950	0.00759	0.00771
	$\hat{f}_{dn}$	0.00339	0.00480	0.00672	0.00451	0.00676	0.00713	0.00910	0.00793	0.00808
	$\hat{f}_d$	0.00419	0.00442	0.00584	0.00507	0.00685	0.00702	0.00674	0.00674	0.00705

CHAPTER 6

Kernel Density and Regression Estimations for Samples with Random Sizes

6.1. Introduction and Notation

In this chapter, on a somewhat independent track, we consider the problem of estimating density and regression functions from samples with random sizes. Such samples arise from many sampling techniques including but not limited to Bernoulli sampling, Poisson sampling and with-replacement sampling with the sample size taken to be the number of distinct units. A key element in studying the properties of any estimator build from random-size samples is the knowledge of the distribution of the random sample size. This distribution could be completely different under different situations. Therefore, developing a general framework under which one can perform kernel density and regression estimations from random-size samples would be difficult. In this chapter, we will confine our attention to density and regression estimations under the case of sampling with-replacement from finite populations. In this case, we have both the full-sample of size n, including all repetitions, and the effective sample which consists of the distinct elements in the full-sample. Clearly, the effective sample size is a random variable. The reason behind considering this specific case is twofold. First, under this case, the distribution of the random sample size is well-studied in literature (e.g., Pathak (1961)). Second, studying density and regression estimators based on distinct units in SRSWR complements the emerging literature on sufficient bootstrapping under which one uses only distinct units in bootstrapping samples rather than full bootstrapping samples (e.g., Singh and Sedory (2011)).

Let the set of labels $U = \{1,\ldots,i,\ldots,N\}$ represent a finite population. Suppose Y is a real-valued random variable taking the values $\mathbf{Y}_U = (Y_1,\ldots,Y_N)$ in the finite population. A subset s is selected form U using SRSWR. The variable Y is observed for the sampled units giving the data $\{y_i; i \in s\}$. Let ν be the effective sample size. That is, ν is the number of distinct units in the sample s. Consequently, ν is a random variable that takes values between 0 and n. Let the values of Y for the set of distinct units in the sample be denoted by $\{y_i^*; i \in s_\nu\}$ where s_ν is the labeling set of the distinct units.

In the literature of survey sampling, several work have been done on investigating the statistical properties of the mean of distinct units from SRSWR, denoted by $\hat{\bar{Y}}_\nu$, as an estimator for the finite population mean $\bar{Y}_U$. Basu (1958) and Raj and Khamis (1958) showed, independently, that $\hat{\bar{Y}}_\nu$ is superior to $\hat{\bar{Y}}_{n,WR}$, the mean of the full-sample in SRSWR. It is also well-known that the sampling variance of $\hat{\bar{Y}}_{n,WR}$ is larger than the sampling variance of $\hat{\bar{Y}}_{n,WOR}$, the sample mean from a size n SRSWOR (e.g., Raj and Khamis (1958)). However, this comparison is not fair because the sampling cost for SRSWR is at most equal to the cost for SRSWOR when both samples are of the same size. Based on the set of distinct units, Pathak (1962) defined a class of estimators for $\bar{Y}_U$. Seth and Rao (1964) held a comparison between $\hat{\bar{Y}}_\nu$ and $\hat{\bar{Y}}_{\nu^*,WOR}$, the sample mean of a SRSWOR of size $\nu^* \equiv E_\nu(\nu)$ where E_ν is the expected value over all possible sets of distinct units. They showed that $\hat{\bar{Y}}_{\nu^*,WOR}$ outperforms $\hat{\bar{Y}}_\nu$. Again, this is not a fair comparison because ν^* might be consistently larger than ν and, hence, the cost is not the same for the two estimators. Rao (1966) compares the efficiency of the last two estimators under equal and unequal sampling schemes and draws similar conclusions to those in Seth and Rao (1964). Joshi (1966) showed that $\hat{\bar{Y}}_\nu$ is an admissible estimator for $\bar{Y}_U$. It should be noted that these results do not automatically apply to kernel density or regression estimators because the smoothing parameters required for such estimators are always dependent on the sample size in the case of IID data (see Section 1.2.2) and on both the sample size and the population size when sampling is done from finite populations (e.g., Bellhouse and Stafford (1999)).

The following notations are used throughout the rest of this chapter. Let $E_{\mathscr{D}}$ denote the expectation with respect to the joint distribution induced by the sampling design and the random sample size ν. The expectation with respect to the marginal distribution induced by the sampling design (the randomization distribution) is denoted by $E_{\mathcal{P}}$. Further, let E_ν and E_ξ be the expectations with respect to the marginal distribution of ν and with respect to the underlying model (superpopulation), respectively. Similar notations are used to refer to the variances; $V_{\mathscr{D}}$, $V_{\mathcal{P}}$, V_ν and V_ξ. The expectation and variance operators with respect to the randomization distribution, the superpopulation model and the distribution of ν jointly are denoted by E_C and V_C, respectively.

The rest of this chapter is organized as follows. In section 6.2, we define a new kernel density estimator based on the set of distinct units from SRSWR. The properties of the proposed estimator are investigated in the same section. Bandwidth selection techniques for the new estimator are discussed in Section 6.3. Section 6.4 summarizes the results of a small simulation study in which we investigate the performance of the proposed density estimator relative to another two standard estimators. Finally, as an application for this new density estimate, nonparametric kernel regression estimation based on the set of distinct units from SRSWR is introduced in section 6.5.

6.2. Proposed Density Estimator and Its Main Properties

Suppose an SRSWR of size n is drawn from a finite population of size N, say U. The values of a study variable Y are observed for the sampled units. We want to estimate the density function of Y, say $f(\cdot)$, using the sample data. Considering only distinct units in the sample and using notations introduced in the previous section, we define the following kernel estimator for $f(y)$:

$$\hat{f}_\nu(y;h) = \frac{1}{\nu h}\sum_{i=1}^{\nu} K\left(\frac{y-y_i^*}{h}\right), \tag{6.1}$$

for all $y \in \mathbb{R}$, where h is a smoothing parameter and K is a kernel function.

We need the following assumptions to derive our main results about the estimator in (6.1).

D.1 *(The density function f)*: The function $f(\cdot)$ is a real-valued function having a bounded second derivative that is continuous and square integrable.

D.2 *(The kernel function K)*: The kernel K satisfies conditions A.2(i–ii) in Section 2.2.2.

D.3 *(The bandwidth h)*: $h_\tau(n_\tau, N_\tau) \equiv h_\tau \equiv h$ is such that $h_\tau \to 0$, $N_\tau h_\tau \to \infty$, $n_\tau h_\tau \to \infty$ as $\tau \to \infty$.

The following theorem gives the MISE of $\hat{f}_\nu(\cdot;h)$ under the combined design-model-based inference framework.

Theorem 6.1. *Suppose Assumptions D.1–D.3 hold. The bias and the MISE of $\hat{f}_\nu(y;h)$, under the combined inference framework, are given by:*

$$Bias_C\left[\hat{f}_\nu(y;h)\right] = \frac{1}{2}h^2 c_K f''(y) + o(h^2) \tag{6.2}$$

and

$$MISE_C\left[\hat{f}_\nu(\cdot;h)\right] = \frac{1}{nh}\left[C^0_{(n,N)} + \frac{n}{N}\right] d_K + \frac{1}{4}h^4 c_K^2 d_{f''} + o\left(h^4 + \frac{1}{Nh}\right), \tag{6.3}$$

where $C^0_{(n,N)} = nN^{-n}\sum_{l=1}^{N-1} l^{n-1}$ and $d_{f''} = \int \{f''(y)\}^2 dy$.

Proof: We prove the bias statement first. Note that

$$E_C\left[\hat{f}_\nu(y)\right] = E_\xi\left[E_\nu\left\{E_{\mathcal{P}}\left[\hat{f}_\nu(y)|\mathbf{Y}_U, \nu\right]|\mathbf{Y}_U\right\}\right], \tag{6.4}$$

with

$$E_{\mathcal{P}}\left[\hat{f}_\nu(y)|\mathbf{Y}_U, \nu\right] = E_{\mathcal{P}}\left[\frac{1}{\nu}\sum_{i=1}^{\nu} K_h(y - y_i^*)\,|\mathbf{Y}_U, \nu\right] = \frac{1}{N}\sum_{i=1}^{N} K_h(y - Y_i), \tag{6.5}$$

where the last equality in (6.5) follows from the fact that given ν, the set of distinct units s_ν is a SRSWOR from U and the sample mean $\nu^{-1}\sum_{i=1}^{\nu} K_h(y - y_i^*)$ is a design-unbiased estimator for the finite population mean of the variable $K_h(y - Y_i)$. Applying the outer expectation in (6.4) to the far right-hand side of (6.5) gives

$$\begin{aligned}
E_C\left[\hat{f}_\nu(y)\right] &= E_\xi\left[\frac{1}{N}\sum_{i=1}^{N} K_h(y-Y_i)\right] \\
&\overset{iid}{=} E_\xi\left[K_h(y-Y_1)\right] \\
&= f(y)+\frac{1}{2}h^2 c_K f''(y)+o(h^2),
\end{aligned} \tag{6.6}$$

where the last equality is a standard result in kernel density estimation (e.g., Wand and Jones (1995, pg. 20)). Note that $E_C\left[\hat{f}_\nu(y)\right]$ does not depend on ν. It depends only on the kernel K and the bandwidth h. The proof of the bias statement is complete upon subtracting $f(y)$ from the right-hand side of (6.6).

Next we evaluate the variance of the estimator $\hat{f}_\nu(y)$ under the combined inference approach as follows. First note that

$$V_C\left[\hat{f}_\nu(y)\right] = E_\xi\left[V_{\mathscr{D}}\left\{\hat{f}_\nu(y)|\mathbf{Y}_U\right\}\right] + V_\xi\left[E_{\mathscr{D}}\left\{\hat{f}_\nu(y)|\mathbf{Y}_U\right\}\right] := H_1 + H_2. \tag{6.7}$$

Using (6.5), we can write

$$H_2 = V_\xi\left[\frac{1}{N}\sum_{i=1}^{N} K_h(y-Y_i)\right] = \frac{1}{Nh}d_K f(y) + o\left(\frac{1}{Nh}\right), \tag{6.8}$$

where the last equality is a standard result in kernel density estimation (e.g., Wand and Jones (1995, pg. 21)). For H_1, we compute the inner variance first.

$$V_{\mathscr{D}}\left\{\hat{f}_\nu(y)|\mathbf{Y}_U\right\} = E_\nu\left\{V_{\mathcal{P}}\left[\hat{f}_\nu(y)|\mathbf{Y}_U,\nu\right]|\mathbf{Y}_U\right\} + V_\nu\left\{E_{\mathcal{P}}\left[\hat{f}_\nu(y)|\mathbf{Y}_U,\nu\right]|\mathbf{Y}_U\right\}. \tag{6.9}$$

Again, using (6.5), the second term in (6.9) is

$$V_\nu\left\{E_{\mathcal{P}}\left[\hat{f}_\nu(y)|\mathbf{Y}_U,\nu\right]|\mathbf{Y}_U\right\} = V_\nu\left[\frac{1}{N}\sum_{i=1}^{N} K_h(y-Y_i)|\mathbf{Y}_U\right] = 0. \tag{6.10}$$

It remains to evaluate the first term in (6.9). First observe that

$$
\begin{aligned}
V_{\mathcal{P}}\left[\hat{f}_{\nu}(y)|\mathbf{Y}_U,\nu\right] =& V_{\mathcal{P}}\left[\frac{1}{\nu}\sum_{i=1}^{\nu}K_h\left(y-y_i^*\right)|\mathbf{Y}_U,\nu\right] \\
=& \left(\frac{1}{\nu}-\frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left[K_h\left(y-Y_i\right)-f_U(y)\right]^2, \qquad (6.11)
\end{aligned}
$$

where the equality in (6.11) follows from Theorem (2.2) in Cochran (1977, pg. 23) because given ν, $\nu^{-1}\sum_{i=1}^{\nu}K_h\left(y-y_i^*\right)$ is a sample mean of a size ν SRSWOR. Therefore,

$$
\begin{aligned}
V_{\mathscr{D}}\left\{\hat{f}_{\nu}(y)|\mathbf{Y}_U\right\} =& E_{\nu}\left\{\left(\frac{1}{\nu}-\frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left[K_h\left(y-Y_i\right)-f_U(y)\right]^2|\mathbf{Y}_U\right\} \\
=& \left\{\left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\frac{1}{N-1}\sum_{i=1}^{N}\left[K_h\left(y-Y_i\right)-f_U(y)\right]^2\right\}. \qquad (6.12)
\end{aligned}
$$

From (6.7) and (6.12) together, we have

$$
\begin{aligned}
H_1 =& \left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\frac{1}{N-1}E_{\xi}\left\{\sum_{i\in U}\left[K_h\left(y-Y_i\right)-f_U(y)\right]^2\right\} \\
=& \left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\frac{1}{N-1}E_{\xi}\left\{\sum_{i\in U}K_h^2\left(y-Y_i\right)-Nf_U^2(y)\right\} \\
=& \left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\frac{1}{N-1}E_{\xi}\left\{\sum_{i\in U}K_h^2\left(y-Y_i\right)-\frac{1}{N}\sum_{\in U}K_h^2\left(y-Y_i\right)\right. \\
&\left.-\frac{1}{N}\sum\sum_{i,j\in U,\ i\neq j}K_h^2\left(y-Y_i\right)\right\} \\
\overset{iid}{=}& \left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\frac{N}{N-1}\left\{E_{\xi}\left[K_h^2\left(y-Y_1\right)\right]-\frac{1}{N}E_{\xi}\left[K_h^2\left(y-Y_1\right)\right]\right\} \\
&-\left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\left\{E_{\xi}\left[K_h\left(y-Y_1\right)\right]\right\}^2 \\
=& \left[E_{\nu}\left(\frac{1}{\nu}\right)-\frac{1}{N}\right]\left\{E_{\xi}\left[K_h^2\left(y-Y_1\right)\right]-\left[E_{\xi}\left\{K_h\left(y-Y_1\right)\right\}\right]^2\right\}. \qquad (6.13)
\end{aligned}
$$

Now, it is clear that we need to know $E_{\nu}\left(\nu^{-1}\right)$. Pathak (1961) derived both exact and approximate formulas for this expectation. The exact formula takes the following form:

$$
E_{\nu}\left(\frac{1}{\nu}\right)=\frac{1}{N^n}\sum_{l=1}^{N}l^{n-1}=\frac{1}{N}+\frac{1}{n}C^0_{(n,N)}, \qquad (6.14)
$$

where $C^0_{(n,N)} = nN^{-n}\sum_{l=1}^{N-1} l^{n-1}$. The first equality in (6.14) was proven in different ways by many authors including Korwar and Serfling (1970) and Lanke (1975). Using (6.14) in (6.13), we get

$$H_1 = \frac{1}{n}C^0_{(n,N)}\left\{E_\xi\left[K_h^2(y-Y_1)\right] - \left[E_\xi\left\{K_h(y-Y_1)\right\}\right]^2\right\}$$
$$= \frac{1}{nh}C^0_{(n,N)}d_K f(y) + o\left(\frac{1}{Nh}\right), \tag{6.15}$$

where the equality in (6.15) follows from standard results in kernel density estimation (e.g., Wand and Jones (1995, pg. 21)).

Now, substitute (6.8) and (6.15) into (6.7) to get

$$V_C\left[\hat{f}_\nu(y)\right] = \frac{1}{nh}\left[C^0_{(n,N)} + \frac{n}{N}\right]d_K f(y) + o\left(\frac{1}{Nh}\right). \tag{6.16}$$

Additionally, adding the squared bias to (6.16) gives the MSE of $\hat{f}_\nu(y)$ under the combined inference framework:

$$MSE_C\left[\hat{f}_\nu(y)\right] = \frac{1}{nh}\left[C^0_{(n,N)} + \frac{n}{N}\right]d_K f(y) + \frac{1}{4}h^4 c_K^2\{f''(y)\}^2 + o\left(h^4 + \frac{1}{Nh}\right). \tag{6.17}$$

Integrating this MSE over y gives the MISE of $\hat{f}_\nu(\cdot)$ and the proof is complete. □

The optimal bandwidth, in the sense that it minimizes the AMISE, takes the following form:

$$h_{opt,\nu_0} = \left[\left(C^0_{(n,N)} + (n/N)\right)d_K/\{c_K^2 d_{f''}\}\right]^{1/5} n^{-1/5}. \tag{6.18}$$

Corollary 6.1. *Suppose, following Korwar and Serfling (1970), we use the approximation:*

$$E_\nu(\nu^{-1}) \approx \left[\frac{1}{n} + \frac{1}{2N} + \frac{(n-1)}{12N^2}\right]. \tag{6.19}$$

Then,

$$MISE_C\left[\hat{f}_\nu(\cdot;h)\right]=\frac{1}{nh}\left[C^1_{(n,N)}+\frac{n}{N}\right]d_K+\frac{1}{4}h^4c_K^2d_{f''}+o\left(h^4+\frac{1}{nh}\right), \tag{6.20}$$

and the optimal bandwidth becomes

$$h_{opt,\nu_1}=\left[\left(C^1_{(n,N)}+(n/N)\right)d_K/\{c_K^2d_{f''}\}\right]^{1/5}n^{-1/5}, \tag{6.21}$$

where $C^1_{(n,N)}=1-\frac{n}{2N}+\frac{n(n-1)}{12N^2}$.

Proof: Korwar and Serfling (1970) argue that the approximation in (6.19) is an excellent approximation which is suitable not only for computations but also for theoretical considerations as long as $3\leq n\leq N$. Using (6.19) in (6.13), we have

$$\begin{aligned}H_1=&\left[\frac{1}{n}-\frac{1}{2N}+\frac{(n-1)}{12N^2}\right]\left\{E_\xi\left[K_h^2(y-Y_1)\right]-\left[E_\xi\left\{K_h(y-Y_1)\right\}\right]^2\right\}\\=&\frac{1}{n}\left[1-\frac{n}{2N}+\frac{n(n-1)}{12N^2}\right]\left\{E_\xi\left[K_h^2(y-Y_1)\right]-\left[E_\xi\left\{K_h(y-Y_1)\right\}\right]^2\right\}\\=&\frac{1}{nh}C^1_{(n,N)}d_Kf(y)+o\left(\frac{1}{nh}\right),\end{aligned} \tag{6.22}$$

Adding (6.8) to (6.22) gives

$$V_C\left[\hat{f}_\nu(y)\right]=\frac{1}{nh}\left[C^1_{(n,N)}+\frac{n}{N}\right]d_Kf(y)+o\left(\frac{1}{nh}\right). \tag{6.23}$$

Finally, adding the squared bias to (6.23) and integrating the result over y gives (6.20). Minimizing the sum of the first two terms on the right-hand side of (6.20) with respect to h gives the optimal bandwidth in (6.21) and the proof is complete. □

Remark: The MISE formula in Corollary 6.1 has an interesting representation. This representation shows that if the population size is very large relative to the sample size such that the sampling fraction $n_\tau/N_\tau\to 0$ as $\tau\to\infty$, then ν would be very close to n because the chances of getting the same unit sampled more than once is very rare and, therefore, we should retrieve the MISE formula that we get when we sample from infinite populations

(IID samples).

Corollary 6.2. *Suppose we use all units in an SRSWR including repetitions to estimate the density $f(y)$. That is, we define the estimator*

$$\hat{f}_{n,WR}(y)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{y-y_i}{h}\right). \tag{6.24}$$

Then,

$$MISE_C\left[\hat{f}_{n,WR}(y)\right]=\frac{1}{nh}\left(1+\frac{n}{N}\right)d_K+\frac{1}{4}h^4c_K^2d_{f''}+o\left(h^4+\frac{1}{nh}\right), \tag{6.25}$$

and the optimal bandwidth is

$$h_{opt,WR_n}=\left[(1+n/N)\,d_K/\{c_K^2d_{f''}\}\right]^{1/5}n^{-1/5}. \tag{6.26}$$

Corollary 6.3. *Suppose we use an SRSWOR of size ν (the number of distinct units in an SRSWR of size n) to estimate the density $f(y)$. That is, we define the estimator*

$$\hat{f}_{\nu,WOR}(y)=\frac{1}{\nu h}\sum_{i=1}^{\nu}K\left(\frac{y-y_i}{h}\right). \tag{6.27}$$

Then,

$$MISE_C\left[\hat{f}_{\nu,WOR}(y)\right]=\frac{1}{\nu h}d_K+\frac{1}{4}h^4c_K^2d_{f''}+o\left(h^4+\frac{1}{\nu h}\right), \tag{6.28}$$

and the optimal bandwidth is

$$h_{opt,WOR_\nu}=\left[d_K/\{c_K^2d_{f''}\}\right]^{1/5}\nu^{-1/5}. \tag{6.29}$$

6.3. Bandwidth Selection

In this section, we address the bandwidth selection problem for the estimator $\hat{f}_\nu(y;h)$ defined in (6.1). It is not hard to note that one can obtain a plug-in bandwidth estimate for $\hat{f}_\nu(y;h)$ through replacing the quantity $d_{f''}$ in (6.18) or in (6.21) by a kernel estimate as discussed in Section 2.5. As we noted in Section 1.2.3, plug-in bandwidth estimators are not completely data-driven estimators because they involve subjective selection of pilot bandwidths. Thus, we focus in this section on cross-validation bandwidth selectors which provide completely data-driven bandwidth estimates. More specifically, we discuss bandwidth selectors for $\hat{f}_\nu(y;h)$ using both least squares cross-validation and biased cross-validation methods.

6.3.1. Least Squares Cross-Validation Method

Note that the model-based $MISE$ of $\hat{f}_\nu(y;h)$ has the form

$$\begin{aligned}MISE\left[\hat{f}_\nu(\cdot;h)\right] &= \int MSE\left[\hat{f}_\nu(y;h)\right]dy\\ &= \int E\left[\hat{f}_\nu(y;h)-f(y)\right]^2 dy\\ &= E\left\{\int\left[\hat{f}_\nu(y;h)\right]^2 dy - 2\int \hat{f}_\nu(y;h)f(y)dy\right\} + \int [f(y)]^2 dy,\end{aligned}$$

where the expectation E is taken with respect to the joint distribution of the Y's and ν. The term $\int [f(y)]^2 dy$ on the right-hand side of the last equality does not depend on h. Thus, minimizing the $MISE$ is equivalent to minimizing the objective function

$$L_\nu(h) = E\left\{\int\left[\hat{f}_\nu(y;h)\right]^2 dy - 2\int \hat{f}_\nu(y;h)f(y)dy\right\}. \tag{6.30}$$

It is readily seen that $L_\nu(h)$ is unknown as it depends on the unknown density f. The idea of cross-validation is to choose h that minimizes an unbiased estimate of $L_\nu(h)$. The following lemma gives an unbiased estimator for $L_\nu(h)$ based on the sect of distinct units in the full-sample.

Lemma 6.1. *Consider the set of distinct units in a size n SRSWR taken from a finite population; $\{y_i^*; i \in s_\nu\}$. An unbiased estimator of $L_\nu(h)$ based on this set of units is given by:*

$$LSCV_\nu(h) = \int \left[\hat{f}_\nu(y;h)\right]^2 dy - 2\nu^{-1} \sum_{i=1}^{\nu} \hat{f}_{-i}(y_i^*;h) \tag{6.31}$$

where $\hat{f}_{-i}(y_i^*;h) = (\nu-1)^{-1} \sum_{j \neq i}^{\nu} K_h(y_i^* - y_j^*)$ is the leave-one-out density estimate.

Proof: Observe that,

$$E\left\{\nu^{-1}\sum_{i=1}^{\nu}\hat{f}_{-i}(Y_i^*)\right\} = E_\nu\left\{E_\xi\left[\nu^{-1}\sum_{i=1}^{\nu}\hat{f}_{-i}(Y_i^*)\big|\nu\right]\right\}.$$

But,

$$\begin{aligned}
E_\xi\left[\nu^{-1}\sum_{i=1}^{\nu}\hat{f}_{-i}(Y_i^*)|\nu\right] &= E_\xi\left[\nu^{-1}(\nu-1)^{-1}\sum_{i=1}^{\nu}\sum_{j\neq i}^{\nu}K_h(Y_i^*-Y_j^*)\big|\nu\right] \\
&\overset{iid}{=} E_\xi\left[K_h(Y_1^*-Y_2^*)\right] \\
&= E_{Y_1^*}\left\{E_{Y_2^*}\left[K_h(Y_1^*-Y_2^*)\right]\right\} \\
&= E_{Y_1^*}\left\{\int K_h(Y_1^*-y)f(y)dy\right\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E\left\{\nu^{-1}\sum_{i=1}^{\nu}\hat{f}_{-i}(Y_i^*)\right\} &= E_\xi\left\{\int K_h(Y_1^*-y)f(y)dy\right\} \\
&= E_\xi\left\{\int \hat{f}_\nu(y)f(y)dy\big|\nu\right\} = E\left\{\int \hat{f}_\nu(y)f(y)dy\right\},
\end{aligned}$$

and the proof is complete. □

According to the least squares cross-validation method, the optimal bandwidth is the value h that minimizes the estimate $LSCV_\nu(h)$ and, consequently, minimizes $L_\nu(h)$. We denote such minimizer by $\hat{h}_{LSCV,\nu}$.

6.3.2. Biased Cross-Validation Method

As an alternative to $\hat{h}_{LSCV,\nu}$, we define a biased cross-validation (BCV) bandwidth selector for $\hat{f}_\nu(y;h)$. From Section 6.2, recall that the *AMISE* of $\hat{f}_\nu(y;h)$ is

$$AMISE\left[\hat{f}_\nu(\cdot;h)\right] = \frac{1}{nh}\left[C^0_{(n,N)} + \frac{n}{N}\right] d_K + \frac{1}{4}h^4 c_K^2 d_{f''}. \tag{6.32}$$

Following the same approach in the IID case with fixed sample size and no replicates, the BCV method starts by estimating the quantity $d_{f''}$ using

$$\hat{d}_{f''} = d_{\hat{f}''_\nu} - \left(\nu h^5\right)^{-1} d_{K''}, \tag{6.33}$$

where $\hat{f}''_\nu(y) = (\nu h^3)^{-1}\sum_{j=1}^{\nu} K''\left(\frac{y-y_j^*}{h}\right)$. Note that

$$\begin{aligned}
d_{\hat{f}''_\nu} =& \int_{\mathbb{R}} \left\{(\nu h^3)^{-1}\sum_{j=1}^{\nu} K''\left(\frac{y-y_j^*}{h}\right)\right\}^2 dy \\
=& (\nu h^3)^{-2}\int_{\mathbb{R}}\left\{\sum_{i=1}^{\nu}\left[K''\left(\frac{y-y_i^*}{h}\right)\right]^2 + 2\sum\sum_{i<j} K''\left(\frac{y-y_i^*}{h}\right)K''\left(\frac{y-y_j^*}{h}\right)\right\} dy \\
=& (\nu h^3)^{-2}\sum_{i=1}^{\nu}\int_{\mathbb{R}}\left[K''\left(\frac{y-y_i^*}{h}\right)\right]^2 dy \\
&+ 2(\nu h^3)^{-2}\sum\sum_{i<j}\int_{\mathbb{R}} K''\left(\frac{y-y_i^*}{h}\right)K''\left(\frac{y-y_j^*}{h}\right) dy \\
=& \nu^{-1}h^{-6}\int_{\mathbb{R}}\left[K''(z)\right]^2 h dz + 2\nu^{-2}h^{-6}\sum\sum_{i<j}\int_{\mathbb{R}} K''(z)K''\left(z+\frac{y_i^*-y_j^*}{h}\right) h dz \\
=& (\nu h^5)^{-1} d_{K''} + 2\nu^{-2}h^{-5}\sum\sum_{i<j}\phi(c_{ij}),
\end{aligned} \tag{6.34}$$

where $\phi(c_{ij}) = \int_{\mathbb{R}} K''(z)K''(z+h^{-1}\{y_i^*-y_j^*\})dz$. Substituting (6.34) into (6.33) gives

$$\hat{d}_{f''} = 2\nu^{-2}h^{-5}\sum\sum_{i<j}\phi(c_{ij}). \tag{6.35}$$

Using (6.35) in (6.32), we can define the BCV objective function as follows:

$$BCV_\nu(h) = \frac{1}{nh}\left[C^0_{(n,N)} + \frac{n}{N}\right] d_K + \frac{1}{2\nu^2 h} c_K^2 \sum\sum_{i<j} \phi(c_{ij}). \tag{6.36}$$

According to the BCV method, the optimal bandwidth is the value of h that minimizes $BCV_\nu(h)$ and, hence, minimizes $AMISE\left[\hat{f}_\nu(\cdot;h)\right]$.

6.4. A Simulation

In this section, we investigate the finite sample properties of the estimator $\hat{f}_\nu(y)$ via a small Monte Carlo study. The performance of this estimator is assessed relative to both the estimator based on the full-sample, $\hat{f}_{n,WR}(y)$ given in (6.24), and the estimator from a SRSWOR of size ν, $\hat{f}_{\nu,WOR}(y)$ in (6.27). We consider sampling from finite populations drawn from three different superpopulations. These populations and other simulation parameters are described in the following subsection.

6.4.1. Simulation Settings

Consider estimating the the density function f of a study variable Y having one of the following distributions: the standard normal distribution $[N(0,1)]$, the lognormal distribution $[lnN(0,0.5)]$ and the mixture normal distribution $[0.5N(-1,2/3)+0.5N(1,2/3)]$. From each of these distributions, we generate a single finite population of size $N = 1,000$ and then draw $m = 10,000$ repeated samples from that finite population. Six sample sizes are considered: $n = 25, 50, 75, 100, 150, 200$. To calculate the first two estimators, $\hat{f}_\nu(y)$ and $\hat{f}_{n,WR}(y)$, we generate repeated samples using the with-replacement simple random sampling scheme. Then, $\hat{f}_\nu(y)$ is calculated for each sample based on the set of distinct elements in that sample while $\hat{f}_{n,WR}(y)$ is computed from all units in the sample. Corresponding to each with-replacement sample, we generate a non-replacement simple random sample of size equal to the number of distinct units in the with-replacement sample and use these samples to compute the third estimator $\hat{f}_{\nu,WOR}(y)$. This set-up makes the three estimators have the same cost as they use exactly same number of distinct units. Each estimator is evaluated at 201 grid points covering the range of Y. Least squares cross-validation bandwidth estimators described in Section 6.3 are used in computing each of the three estimators. We used the R function *bw.ucv* in the package *stats* to compute the cross-validation bandwidth estimates. The standard normal kernel is used for all estimators.

6.4.2. Simulation Results

This section reports the results of the simulation study described above. Table 6.1 gives the MISE values for each of the three estimators under each of the three distributions and each sample size. Each MISE value is an area under the MSE curve. These MSE curves are depicted in Figures 6.1–6.3. From these results, we notice that in all cases, the two estimators that do not contain any replicated sample values, $\hat{f}_V(y)$ and $\hat{f}_{V,WOR}(y)$, outperform the estimator $\hat{f}_{n,WR}(y)$ which uses all n units from a with-replacement simple random sample. This fact becomes clearer as the sample size increases because this in turn increases the number of replicates within the with-replacement sample. The two estimators $\hat{f}_V(y)$ and $\hat{f}_{V,WOR}(y)$ perform closely in almost all cases and there is no ultimate winner among them. These results suggest that if the sample data contain repetitions that may result from the sampling mechanism or the nature of the phenomenon generating the data, these repetitions should be removed from the data before performing kernel density estimation because not only they add no extra information, they also cause the resulting estimators to be less accurate.

Table 6.1: MISE values for three KDEs under three distributions and six sample sizes.

n	*NORMAL*			*MIXNORMAL*			*LOGNORMAL*		
	$\hat{f}_{n,WR}$	$\hat{f}_V$	$\hat{f}_{V,WOR}$	$\hat{f}_{n,WR}$	$\hat{f}_V$	$\hat{f}_{V,WOR}$	$\hat{f}_{n,WR}$	$\hat{f}_V$	$\hat{f}_{V,WOR}$
25	0.02627	0.02484	0.02459	0.02909	0.02785	0.02798	0.06923	0.06636	0.06538
50	0.01542	0.01409	0.01405	0.01964	0.01817	0.01799	0.04225	0.03950	0.03955
75	0.01171	0.01027	0.01013	0.01595	0.01447	0.01452	0.03295	0.02997	0.02969
100	0.00945	0.00794	0.00800	0.01413	0.01226	0.01221	0.02726	0.02404	0.02451
150	0.00751	0.00585	0.00573	0.01175	0.00978	0.00983	0.02215	0.01827	0.01850
200	0.00623	0.00452	0.00455	0.01085	0.00840	0.00845	0.01969	0.01517	0.01517

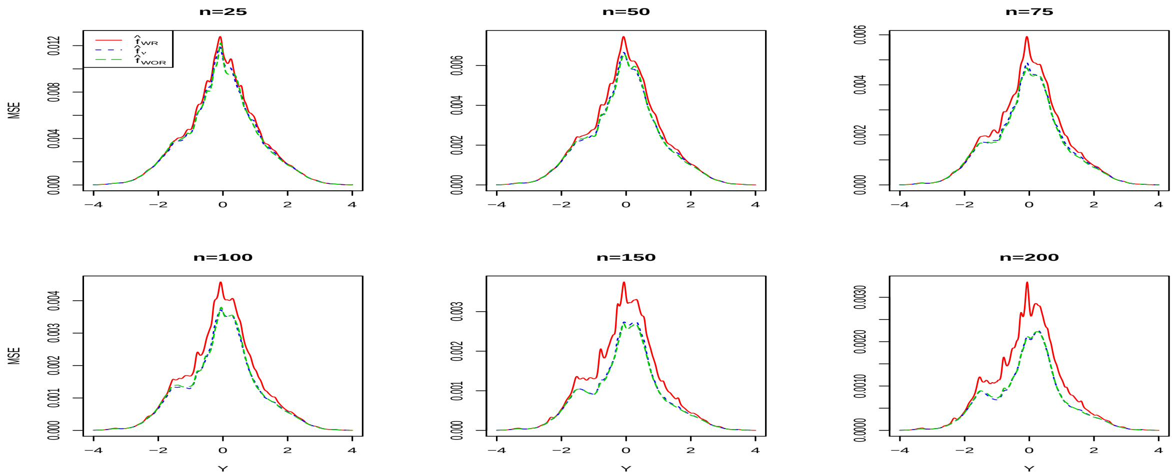

Figure 6.1: MSE for three KDEs under the standard normal distribution.

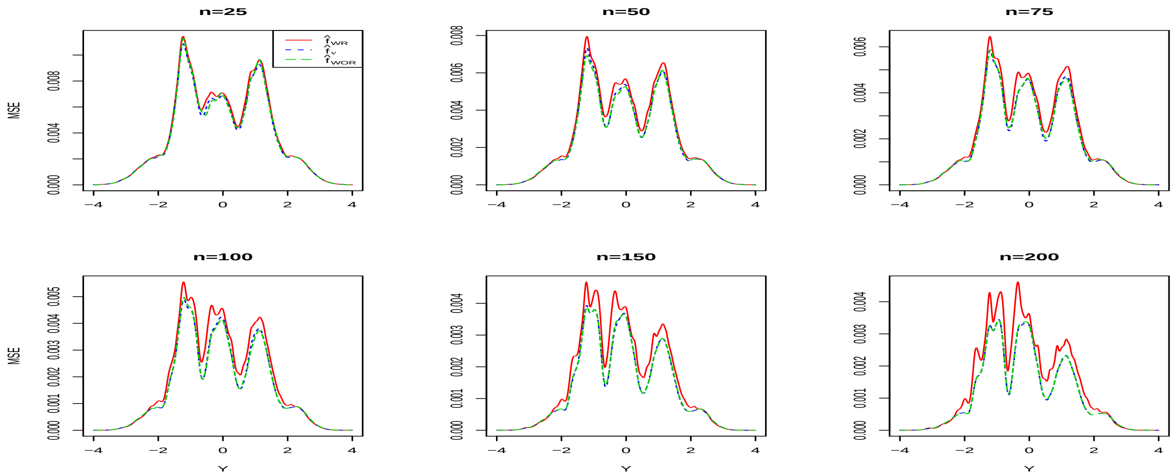

Figure 6.2: MSE for three KDEs under the mixture normal distribution.

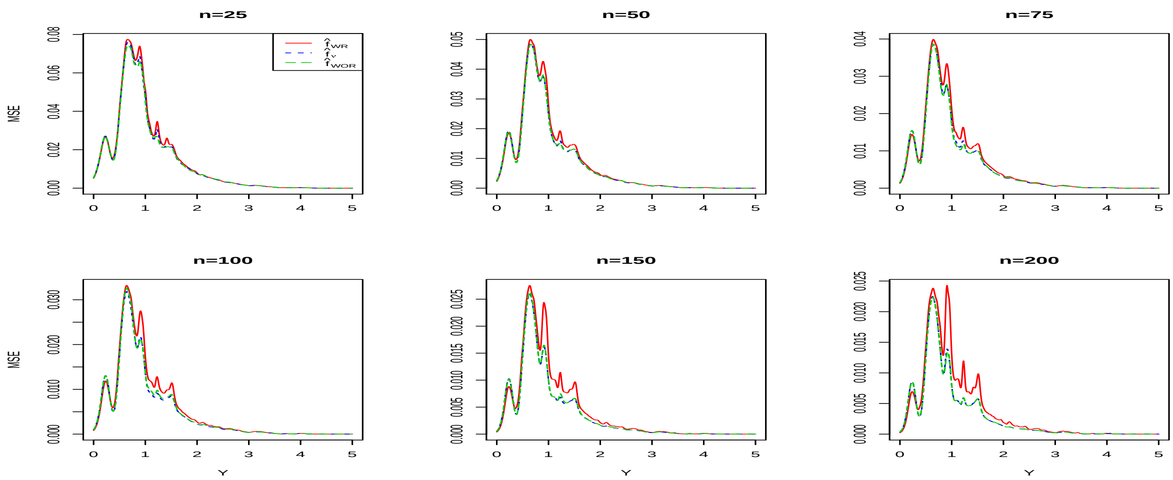

Figure 6.3: MSE for three KDEs under the lognormal distribution.

6.5. Application: Regression Estimation

In this section, as an application for the density estimator $\hat{f}_V(y)$, we study the behavior of nonparametric kernel regression estimators based on the set of distinct units in simple random samples drawn with replacement from finite populations. Local polynomial regression estimators are well-established in literature. In this section, our focus is put on local constant regression estimators, i.e., Nadaraya-Watson-type kernel regression estimators based on the distinct units in the sample.

6.5.1. Proposed Regression Estimator

Consider the random design regression model where both response and independent variables are random variables. Here, we assume that we have a single independent variable X and a single response variable Y. We observe a bivariate sample $(X_1,Y_1),\cdots,(X_n,Y_n)$ of the random pair (X,Y) which has a joint continuous density function $t(x,y)$ with marginals $g_X(x)$ and $f_Y(y)$. The regression model describing the relationship between the two variables can be written as follows:

$$Y_i{=}m(X_i)+\sigma(X_i)\varepsilon_i, \quad i=1,2,\cdots,N,$$

where $m(x)=E_\xi(Y|X=x)$ and $\sigma^2(x)=V_\xi(Y|X=x)$ are unspecified smooth functions and the variables ε_i are independent with zero mean and unit variance. It is not hard to notice that the regression function $m(x)$ can be written as follows:

$$m(x){=}\int yt(y|x)dy=\frac{\int yt(x,y)dy}{g(x)}. \tag{6.37}$$

To obtain a kernel estimator for $m(x)$, one needs to replace both quantities in the numerator and the denominator of (6.37) by kernel estimates. Assuming the same setting of simple random sampling with replacement as in Section 6.1 and adapting similar notation used in the same section while keeping in mind that our sample is now consisted of pairs (x,y) rather than single values of Y, we propose the following Nadaraya-Watson-type kernel regression estimator for $m(x)$ based on the set of distinct pairs in the sample $\{(x_i^*, y_i^*); i \in s_\nu\}$:

$$\hat{m}_\nu(x;b) = \frac{\nu^{-1}\sum_{i=1}^{\nu} y_i^* K_b(x_i^* - x)}{\hat{g}_\nu(x)}, \tag{6.38}$$

where $\hat{g}_\nu(x) = \nu^{-1}\sum_{i=1}^{\nu} K_b(x_i^* - x)$, $K_h(\cdot) = b^{-1}K(\cdot/b)$, b is a bandwidth and K is a kernel function.

In the following subsection, we derive the formula for the asymptotic MSE of the estimator $\hat{m}_\nu(x;b)$. This formula serves as a measure of the efficiency of the estimator and as the basis for selecting the optimal amount of smoothing needed for the estimator.

6.5.2. MSE Approximation

We use the following assumptions to prove Theorem 6.2 which gives the MSE of $\hat{m}_\nu(x;b)$ under the combined design-model-based inference framework.

E.1 The functions $m''(\cdot)$ and $\sigma(\cdot)$ are continuous on the interval $[0,1]$.

E.2 The density function $g_X(\cdot)$ is supported on $[0,1]$ and $g'_X(\cdot)$ is continuous.

E.3 The kernel function $K(\cdot)$ is symmetric about zero and supported on $[-1,1]$.

E.4 $b_\tau(n_\tau, N_\tau) \equiv b_\tau \equiv b$ is such that $b_\tau \to 0$, $N_\tau b_\tau \to \infty$, $n_\tau b_\tau \to \infty$ as $\tau \to \infty$.

E.5 The estimation point x is such that $b < x < 1 - b$ for all sufficiently large n.

Theorem 6.2. *Suppose Assumptions E.1–E.5 hold. The MSE of the estimator $\hat{m}_\nu(x;b)$ under the combined inference framework is given by:*

$$MSE_C[\hat{m}_\nu(x;b)] = \frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right] d_K \frac{\sigma^2(x)}{g(x)} + \frac{1}{4} b^4 c_K^2 \left[m''(x) + \frac{2m'(x)g'(x)}{g(x)}\right]^2 + o\left(b^4 + \frac{1}{Nb}\right). \tag{6.39}$$

Proof: We separately approximate each of the two components of the MSE of $\hat{m}_\nu(x;b)$; bias and variance, as follows. First, observe that $\hat{m}_\nu(x;b)$ is a ratio of the two quantities

$$Q := \frac{1}{\nu}\sum_{i=1}^{\nu} y_i^* K_b(x - x_i^*) \tag{6.40}$$

and

$$W := \hat{g}_\nu(x) = \frac{1}{\nu}\sum_{i=1}^{\nu} K_b(x - x_i^*). \tag{6.41}$$

Thus, approximating the expectation and the variance of $\hat{m}_\nu(x;b)$ requires the moments of U and W. In the following, we work on the derivation of these moments. Note that, from Section 6.2, we have

$$E_C(W) = g(x) + \frac{1}{2}b^2 g''(x)c_K + o(b^2) \tag{6.42}$$

and

$$V_C(W) = \frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right] g(x)d_K + o\left(\frac{1}{Nb}\right). \tag{6.43}$$

On the other hand,

$$\begin{aligned}
E_C(Q) &= E_\xi\left[E_\nu\left\{E_{\mathcal{P}}\left[\frac{1}{\nu}\sum_{i=1}^{\nu} Y_i^* K_b(x - X_i^*)|\mathbf{X}_U, \mathbf{Y}_U, \nu\right]|\mathbf{X}_U, \mathbf{Y}_U\right\}\right]\\
&= E_\xi\left[E_\nu\left\{\frac{1}{N}\sum_{i=1}^{N} Y_i K_b(x - Y_i)|\mathbf{X}_U, \mathbf{Y}_U\right\}\right]\\
&= E_\xi\left[\frac{1}{N}\sum_{i=1}^{N} Y_i K_b(x - Y_i)\right]\\
&= E_\xi\left[Y_1 K_b(x - Y_1)\right]\\
&= \iint_{\mathbb{R}} y_1 K_b(x - x_1)t(y_1|x_1)g(x_1)dy_1dx_1\\
&= \int_{\mathbb{R}} m(x_1)K_b(x - x_1)g(x_1)dx_1\\
&= \int_{\mathbb{R}} m(x + bu)K(u)g(x + bu)du\\
&\approx \int_{\mathbb{R}} K(u)[m(x) + bu\, m'(x) + \frac{1}{2}b^2u^2m''(x)][g(x) + bu\, g'(x)]du
\end{aligned} \tag{6.44}$$

$$
\begin{aligned}
&=m(x)g(x)+b^2m'(x)g'(x)c_K+\frac{1}{2}b^2m''(x)g(x)c_K+o(b^2)\\
&=m(x)g(x)+b^2[m'(x)g'(x)+\frac{1}{2}m''(x)g(x)]c_K+o(b^2),
\end{aligned}
\tag{6.45}
$$

and

$$
V_C(Q)=E_\xi\left[V_D\left\{Q|\mathbf{X}_U,\mathbf{Y}_U\right\}\right]+V_\xi\left[E_D\left\{Q|\mathbf{X}_U,\mathbf{Y}_U\right\}\right]:=A_1+A_2. \tag{6.46}
$$

Using (6.44),

$$
\begin{aligned}
A_2&=\frac{1}{N}V_\xi\left[Y_1K_b(x-X_1)\right]\\
&=\frac{1}{N}E_\xi\left[Y_1K_b(x-X_1)\right]^2-\frac{1}{N}E_\xi^2\left[Y_1K_b(x-X_1)\right]\\
&:=A_{21}-A_{22}.
\end{aligned}
\tag{6.47}
$$

Observe that,

$$
\begin{aligned}
A_{21}&=\frac{1}{N}\iint_{\mathbb{R}}y_1^2K_b^2(x-x_1)t(y_1|x_1)g(x_1)dy_1dx_1\\
&=\frac{1}{N}\int_{\mathbb{R}}K_b^2(x-x_1)g(x_1)\left[\int_{\mathbb{R}}y_1^2t(y_1|x_1)dy_1\right]dx_1\\
&=\frac{1}{N}\int_{\mathbb{R}}\left[\sigma^2(x_1)+m^2(x_1)\right]K_b^2(x-x_1)g(x_1)dx_1\\
&=\frac{1}{Nb}\int_{\mathbb{R}}\left[\sigma^2(x+bu)+m^2(x+bu)\right]K^2(u)g(x+bu)du\\
&=\frac{1}{Nb}\left[\sigma^2(x)g(x)+m^2(x)g(x)\right]d_K+o\left(\frac{1}{Nb}\right).
\end{aligned}
\tag{6.48}
$$

From (6.44),

$$
A_{22}=\frac{1}{N}\left[m(x)g(x)+O(b^2)\right]^2. \tag{6.49}
$$

Therefore,

$$
A_2=\frac{1}{Nb}\left[\sigma^2(x)g(x)+m^2(x)g(x)\right]d_K-\frac{1}{N}\left[m(x)g(x)+O(b^2)\right]^2+o\left(\frac{1}{Nb}\right)
$$

$$= \frac{1}{Nb}\left[\sigma^2(x)g(x) + m^2(x)g(x)\right] d_K + o\left(\frac{1}{Nb}\right). \tag{6.50}$$

Moreover,

$$\begin{aligned} V_{\mathcal{P}}\left[Q|\mathbf{X}_U, \mathbf{Y}_U, \nu\right] =& V_{\mathcal{P}}\left[\frac{1}{\nu}\sum_{i=1}^{\nu} y_i^* K_b(x - x_i^*)|\mathbf{X}_U, \mathbf{Y}_U, \nu\right] \\ &= \left(\frac{1}{\nu} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}\left[Y_i K_b(x - X_i) - \frac{1}{N}\sum_{i=1}^{N} Y_i K_b(x - X_i)\right]^2, \end{aligned}$$

and, hence,

$$\begin{aligned} V_D\left[Q|\mathbf{X}_U, \mathbf{Y}_U\right] =& \left[E_\nu\left(\frac{1}{\nu}\right) - \frac{1}{N}\right]\frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 K_b^2(x - X_i) - N\left\{\frac{1}{N}\sum_{i=1}^{N} Y_i K_b(x - X_i)\right\}^2\right] \\ &= \left[E_\nu\left(\frac{1}{\nu}\right) - \frac{1}{N}\right]\frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 K_b^2(x - X_i) - \frac{1}{N}\sum_{i=1}^{N} Y_i^2 K_b^2(x - X_i)\right. \\ &\qquad \left. - \frac{1}{N}\sum\sum_{i \neq j} Y_i Y_j K_b(x - X_i) K_b(x - X_j)\right]. \end{aligned}$$

Consequently,

$$\begin{aligned} A_1 =& \left[E_\nu\left(\frac{1}{\nu}\right) - \frac{1}{N}\right]\frac{N}{N-1}\left[E_\xi\left\{Y_1^2 K_b^2(x - X_1)\right\} - \frac{1}{N}E_\xi\left\{Y_1^2 K_b^2(x - X_1)\right\}\right] \\ &- \left[E_\nu\left(\frac{1}{\nu}\right) - \frac{1}{N}\right] E_\xi\left[Y_1 Y_2 K_b(x - X_1) K_b(x - X_2)\right] \\ =& \left[E_\nu\left(\frac{1}{\nu}\right) - \frac{1}{N}\right]\left[E_\xi\left\{Y_1^2 K_b^2(x - X_1)\right\} - E_\xi^2\left\{Y_1 K_b(x - X_1)\right\}\right] \\ =& \frac{1}{nb} C^0_{(n,N)}\left[\sigma^2(x) + m^2(x)\right] g(x) d_K + o\left(\frac{1}{Nb}\right), \end{aligned}$$

and, thus,

$$V_C(Q) = \frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right]\left[\sigma^2(x) + m^2(x)\right] g(x) d_K + o\left(\frac{1}{Nb}\right). \tag{6.51}$$

Next, observe that

$$
\begin{aligned}
E_C(QW) &= E_C\left[\left\{\frac{1}{\nu}\sum_{i=1}^{\nu} y_i^* K_b(x-x_i^*)\right\}\left\{\frac{1}{\nu}\sum_{i=1}^{\nu} K_b(x-x_i^*)\right\}\right] \\
&= E_C\left[\frac{1}{\nu^2}\sum_{i=1}^{\nu} y_i^* K_b^2(x-x_i^*)\right] + E_C\left[\frac{1}{\nu^2}\sum\sum_{i\neq j} y_i^* K_b(x-x_i^*)K_b(x-x_j^*)\right] \\
&:= B_1 + B_2. \qquad (6.52)
\end{aligned}
$$

First,

$$
\begin{aligned}
B_1 &= E_\nu\left(\frac{1}{\nu}\right) E_\xi\left[Y_1 K_b^2(x-X_1)\right] \\
&= E_\nu\left(\frac{1}{\nu}\right)\int_{\mathbb{R}} m(x_1)K_b^2(x-x_1)g(x_1)dx_1 \\
&= E_\nu\left(\frac{1}{\nu}\right)\left[\frac{1}{b}m(x)g(x)d_K + o\left(\frac{1}{Nb}\right)\right] \\
&= \frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right] m(x)g(x)d_K + o\left(\frac{1}{Nb}\right).
\end{aligned}
$$

Second,

$$
\begin{aligned}
B_2 &= E_\xi\left[E_\nu\left\{E_{\mathcal{P}}\left[\frac{1}{\nu^2}\sum\sum_{i,j\in U, i\neq j} I_i I_j Y_i K_b(x-X_i)K_b(x-X_j)|\mathbf{X}_U, \mathbf{Y}_U, \nu\right]|\mathbf{X}_U, \mathbf{Y}_U\right\}\right] \\
&= E_\xi\left[E_\nu\left\{\frac{1}{\nu^2}\sum\sum_{i,j\in U, i\neq j} \pi_{ij} Y_i K_b(x-X_i)K_b(x-X_j)|\mathbf{X}_U, \mathbf{Y}_U\right\}\right] \\
&= E_\xi\left[E_\nu\left\{\frac{\nu(\nu-1)}{\nu^2 N(N-1)}\right\}\sum\sum_{i,j\in U, i\neq j} Y_i K_b(x-X_i)K_b(x-X_j)\right] \\
&= \left\{1 - E_\nu\left(\frac{1}{\nu}\right)\right\}\frac{1}{N(N-1)}\sum\sum_{i,j\in U, i\neq j} E_\xi\left[Y_i K_b(x-X_i)\right]E_\xi\left[K_b(x-X_j)\right] \\
&= \left\{1 - E_\nu\left(\frac{1}{\nu}\right)\right\}\left[m(x)g(x) + b^2\left\{m'(x)g'(x) + \frac{1}{2}m''(x)g(x)\right\}c_K + o(b^2)\right] \\
&\quad \left[g(x) + \frac{1}{2}b^2 g''(x)c_K + o(b^2)\right] \\
&= \left\{1 - \frac{1}{n}\left(C^0_{(n,N)} + \frac{n}{N}\right)\right\}\left[m(x)g^2(x) + \frac{1}{2}b^2\left\{m(x)g''(x)c_K + 2m'(x)g'(x)\right.\right. \\
&\qquad\qquad \left.\left. + m''(x)g(x)\right\}g(x)c_K + o(b^2)\right] \\
&= m(x)g^2(x) + \frac{b^2}{2}\left[m(x)g''(x)c_K + 2m'(x)g'(x) + m''(x)g(x)\right]g(x)c_K + o\left(b^2 + \frac{1}{n}\right).
\end{aligned}
$$

Substituting these results into (6.52), we get

$$E_C(QW) = m(x)g^2(x) + \frac{1}{2}b^2\left\{m(x)g''(x)c_K + 2m'(x)g'(x) + m''(x)g(x)\right\}g(x)c_K$$
$$+\frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right]m(x)g(x)d_K + o\left(b^2 + \frac{1}{Nb}\right). \tag{6.53}$$

Using (6.42) and (6.44), we can write

$$E_C(Q)E_C(W) = m(x)g^2(x) + \frac{1}{2}b^2\left[m(x)g''(x) + 2m'(x)g'(x) + m''(x)g(x)\right]g(x)c_K$$
$$+o(b^2). \tag{6.54}$$

From (6.53) and (6.54), we have

$$Cov_C(Q,W) = \frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right]m(x)g(x)d_K + o\left(\frac{1}{Nb}\right). \tag{6.55}$$

Now, using (6.42) and (6.44), we have the following first order approximation of the expectation of $\hat{m}_V(x;b)$:

$$E_C[\hat{m}_V(x;b)] = m(x) + \frac{1}{2}b^2\left[m''(x) + \frac{2m'(x)g'(x)}{g(x)}\right]c_K + o(b^2). \tag{6.56}$$

Additionally, since $\hat{m}_V(x;b) = U/W$, we have

$$V_C[\hat{m}_V(x;b)] \approx \left[\frac{E_C(U)}{E_C(W)}\right]^2\left[\frac{V_C(U)}{E_C^2(U)} + \frac{V_C(W)}{E_C^2(W)} - 2\frac{Cov_C(U,W)}{E_C(U)E_C(W)}\right]. \tag{6.57}$$

Using the moments of U and W, given above, to substitute in (6.57), we have the following first order approximation of the variance of $\hat{m}_V(x;b)$:

$$V_C[\hat{m}_V(x;b)] = \frac{1}{nb}\left[C^0_{(n,N)} + \frac{n}{N}\right]\frac{\sigma^2(x)}{g(x)}d_K + o\left(\frac{1}{Nb}\right). \tag{6.58}$$

The proof of the theorem follows from adding the square of the last two terms on the right-hand side of (6.56) to the right-hand side of (6.58). □

CHAPTER 7

Conclusions and Future Research

7.1. Conclusions

The concept of probability density functions is a vital concept in statistics. Density function estimates represent a corner stone in a wide range of statistical analyses. For instance, density estimates can be used for exploratory data analysis and model diagnostics. Additionally, these estimates are very useful for conducting statistical inference such as discriminant and cluster analyses, hypothesis testing and estimation of density functionals. Thus, enormous literature has been directed to the problem of estimating density functions. The existing literature on density function estimation can be categorized into two broad themes: parametric and nonparametric methods. Among the nonparametric techniques, the kernel method provides a simple and efficient way for estimating density functions. The vast majority of the work in kernel density estimation was conducted under the assumption of having IID samples. Very limited contributions were directed to the case of sampling from finite populations, i.e., complex survey data (see Section 1.4 for a review of these contributions). To the best of our knowledge, there does not exist any literature on the use of auxiliary information in kernel density estimation from complex survey data.

In this dissertation, we studied three new kernel density estimators that use univariate auxiliary information in the context of complex surveys. The new estimators are shown to perform very well relative to standard estimators that ignore the auxiliary information specially when the relationship between a relevant auxiliary variable and the study variable is modeled correctly. A general recommendation that comes out of our work in Chapters 2–4 says the following: "*when considering kernel density estimation from complex survey data, if relevant auxiliary data is available, it is worthwhile to incorporate it in the estimation process to increase the efficiency of the resulting density estimators.*" By relevant auxiliary data, we mean any auxiliary data that is significantly correlated with the study variable.

In Chapter 6, we studied the problem of estimating density and regression functions using the kernel method when the effective sample size is a random variable. The specific situation of with-replacement sampling was considered. Under this situation, we derived asymptotic expressions for the bias, variance and MSE of both kernel density and regression estimators. In a small Monte Carlo study, the KDE based on the set of distinct units (for which the sample size is a random variable) was found superior to the KDE based on all sample units and competitive to the KDE based on a without-replacement sample of size equal to the number of distinct units in the with-replacement sample.

7.2. Future Research

Our work in this dissertation opens many future research problems. Some specific points that we are planning to consider in the near future include:

(i) Extending the estimators in Chapters 2–4 to the setting of multivariate auxiliary data. Under this setting, we can use multivariate parametric regression models to model the relationship between a vector of auxiliary variables and the study variable. Then, we use fitted values obtained from these models to produce kernel density estimators analogous to the model-assisted estimator we proposed in Chapter 2. If the form of the relationship between the study variable and the vector of auxiliary variables is unknown, we suggest using generalized additive models to obtain the fitted values needed for constructing model-assisted density estimators similar to the one we proposed in Chapter 3.

(ii) Another promising point is the application of the model-calibrated pseudo empirical likelihood approach, due to Wu and Sitter (2001), to efficiently incorporate available complete auxiliary information into kernel density estimates.

(iii) We are also interested in investigating how cross-validation methods can be used to obtain data-driven estimates of the smoothing parameters in the three proposed density estimators of Chapters 2–4.

(iv) Investigating how auxiliary data can be used in conjunction with other density estimation techniques, such as the orthogonal series technique, is another point that we plan to consider in our future research.

(v) We would also consider investigating the finite sample properties of the regression estimator we proposed in Section 6.5.

(vi) Finally, we would consider revisiting the problem of estimating density and regression functions using the kernel method under situations where the sample size is random other than the case of with-replacement sampling.

Bibliography

Abadir, K. M. and Lawford, S. (2004). Optimal asymmetric kernels. *Economics Letters*, 83:61–68.

Ahmad, I. A. (1981). A note on nonparametric density estimation for dependent variables using delta method. *The Annals Institute of Statistics and Mathematics*, 33:247–254.

Ahmad, I. A. (2002). Kernel estimation in a continuous randomized response model. *Handbook of applied econometrics and statistical inference*, 165:97–114.

Ahmad, I. A. and Ran, I. S. (2004). Data based bandwidth selection in kernel density estimation with parametric start via kernel contrasts. *Journal of Nonparametric Statistics*, 16:841–877.

Basu, D. (1958). On sampling with and without replacement. *Sankhya*, 20:287–294.

Bellhouse, D. R. and Stafford, J. E. (1999). Density estimation from complex surveys. *Statistica Sinica*, 9:407–424.

Bleuer, S. R. and Kratina, I. S. (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics*, 33:2789–2810.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford University Press, New York.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71:353–360.

Breidt, F. J., Claeskens, G., and Opsomer, J. D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika*, 92:831–846.

Breidt, J. F. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28:1026–1053.

Breunig, R. V. (2001). Density estimation for clustered data. *Econometric Reviews*, 20:353–367.

Breunig, R. V. (2008). Nonparametric density estimation for stratified samples. *Statistics and Probability Letters*, 78:2194–2200.

Buskirk, T. D. (1999). *Using nonparametric methods for density estimation with complex survey data*. PhD thesis, Department of Mathematics, Arizona State University.

Buskirk, T. D. and Lohr, S. L. (2005). Asymptotic properties of kernel density estimation with complex survey data. *Journal of Statistical Planning and Inference*, 128:165–190.

Chambers, R. L., Dorfman, A. H., and Hall, P. (1992). Properties of estimators of the finite population distribution function. *Biometrika*, 79:577–582.

Chambers, R. L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73:597–604.

Chaudhuri, A. and Stenger, H. (2005). *Survey Sampling: Theory and Methods*. Chapman and Hall/CRC, USA.

Chen, J. and Wu, C. (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistics Sinica*, 12:1223–1239.

Chen, S. X. (2000). Probability density function estimation using gamma kernels. *The Annals Institute of Statistics and Mathematics*, 52:471–480.

Cochran, W. G. (1977). *Sampling Techniques*. Wiley, New York.

Dassanayake, D. P. A., Volobouev, I., and Trindade, A. A. (2015). Local orthogonal polynomial expansion for density estimation. *arXiv:1505.00275v1 [stat.AP]*.

Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.

Dorfman, A. H. (1992). Non-parametric regression for estimating totals in finite populations. In *ASA Proceedings of the Section on Survey Research Methods. American Statistical Association (Alexandria, VA)*, pages 622–625.

Dorfman, A. H. and Hall, P. (1993). Estimators of the finite population distribution function using nonparametric regression. *The Annals of Statistics*, 21:1452–1475.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society, Ser. B*, 57:371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, New York.

Fuller, W. A. (2009). *Sampling Statistics*. Wiley, New Jersey.

Glad, I. K., Hjort, N. L., and Ushakov, N. G. (2003). Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, 30:415–427.

Hájek, J. (1960). Limiting distribuions in simple random sampling from a finite population. *Publications of Mathematical Institute of Hungarian Academy of Sciences, Ser. A*, 5:361–374.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Journal of Statistics and Probability Letters*, 6:109–115.

Harms, T. and Duchesne, P. (2010). On kernel nonparametric regression designed for complex survey data. *Metrika*, 72:111–138.

Hartley, H. O. and Sielken, R. L. (1975). A "super-population viewpoint" for finite population sampling. *Biometrics*, 31:411–422.

Heidenreich, N. B., Schindler, A., and Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis*, 97:403–433.

Hengartner, N. W., Wegkamp, M. H., and Løber, E. M. (2002). Bandwidth selection for local linear regression smoothers. *Journal of the Royal Statistical Society*, 64:791–804.

Howell, K. B. (2001). *Principles of Fourier Analysis*. Chapman & Hall /CRC Press, New York.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *JASA*, 77:89–96.

Johnson, A. A., Breidt, F. J., and Opsomer, J. D. (2008). Estimating distribution functions from survey data using nonparametric regression. *Journal of Statistical Theory and Practice*, 2:419–431.

Joshi, V. M. (1966). Admissibility of the sample mean as estimate of the mean of a finite population. *The Annals of Mathematical Statistics*, 39:606–620.

Korwar, R. M. and Serfling, R. J. (1970). On averaging over distinct units in sampling with replacement. *The Annals of Mathematical Statistics*, 41:2132–2134.

Lanke, J. (1975). *Some contributions to the theory of survey sampling*. PhD thesis, Department of Mathematical Statistics, University of Lund, Sweden.

Li, Q. and Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, USA.

Montanari, G. E. and Ranalli, M. G. (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association*, 100:1429–1442.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9:141–142.

Opsomer, J. D., Breidt, F. J., Moisen, G. G., and Kauermann, G. (2007). Model-assisted estimation of forest resources with generalized additive models. *Journal of the American Statistical Association*, 102:400–416.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press, New York.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076.

Pathak, P. K. (1961). On the evaluation of moments of distinct units in a sample. *Sankhya, Ser. A*, 23:415–420.

Pathak, P. K. (1962). On simple random sampling with replacement. *Sankhya, Ser. A*, 24:287–302.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61:317–337.

Pons, O. (2011). *Functional Estimation for Density, Regression and Processes*. World Scientific Publishing Co. Pte. Ltd., Singapore.

R Development Core Team (2011). R: a language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing.*, ISBN: 3-90005:1–70–0.

Raj, D. and Khamis, S. H. (1958). Some remarks on sampling with replacement. *The Annals of Mathematical Statistics*, 39:550–557.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics*, 10:462–474.

Rao, J. K. N. (1966). On the comparison of sampling with and without replacement. *Review of the International Statistical Institute*, 34:125–138.

Rao, J. N. K., Kovar, J. G., and Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77:365–375.

Robinson, P. M. and Särndal, C. E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya, Ser. B*, 45:240–248.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27:832–837.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9:65–78.

Rueda, M., Martinez, S., Martinez, H., and Arcos, A. (2007). Estimation of the distribution function with calibration methods. *Journal of Statistical Planning and Inference*, 137:435–448.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *JASA*, 90:1257–1270.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Scott, D. W. (2004). Multivariate density estimation and visualization. *Papers / Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE)*, 16:1–23.

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.

Scott, D. W. and Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82:1131–1146.

Sen, P. K. (1988). Asymptotics in finite population sampling. *Handbook of Statistics*, 6:291–331.

Seth, G. R. and Rao, J. K. N. (1964). On the comparison between simple random sampling with and without replacement. *Sankhy, Ser. A*, 26:85–86.

Sheather, S. J. (2004). Density estimation. *Statistical Science*, 9:588–597.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of The Royal Statistical Society, Ser. B*, 53:683–690.

Silva, N. and Skinner, C. J. (1995). Estimating distribution functions with auxiliary information using poststratification. *Journal of Official Statistics*, 11:249–277.

Simonoff, J. (1996). *Smoothing Methods in Statistics*. Springer, London.

Singh, S. and Sedory, S. A. (2011). Sufficient bootstrapping. *Computational Statistics and Data Analysis*, 55:1629–1637.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, 12:1285–1297.

Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association*, 85:470–477.

Terrell, G. R. and Scott, D. W. (1980). On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, 8(5):1160–1163.

Terrell, G. R. and Scott, D. W. (1985). Oversmoothed density estimates. *Journal of the American Statistical Association*, 80:209–214.

Thompson, M. E. (1997). *Theory of Sample Surveys*. Chapman and Hall, London.

Tran, C. J. (1990). Kernel density estimation under dependence. *Statistics and Probability Letters*, 10:193–201.

Wand, M. and Jones, M. (1995). *Kernel Smoothing*. Chapman and Hall, London.

Watson, G. S. (1964). Smooth regression analysis. *Sankya, Series A*, 26:359–372.

Wu, C. (1999). *The effective use of complete auxiliary information from survey data*. PhD thesis, Department of Mathematics and Statistics, Simon Fraser University.

Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96:185–193.

VITA

Sayed A. Mostafa

Candidate for the Degree of

Doctor of Philosophy

Dissertation: NONPARAMETRIC KERNEL DENSITY ESTIMATION USING AUXILIARY INFORMATION FROM COMPLEX SURVEY DATA

Major Field: Statistics

Biographical:

- Education:
  - Completed the requirements for the Doctor of Philosophy in Statistics at Oklahoma State University, Stillwater, Oklahoma in July, 2016.
  - Completed the requirements for the Master of Science in Statistics at Oklahoma State University, Stillwater, Oklahoma in May, 2014.
  - Completed the requirements for the Bachelor of Science in Statistics at Cairo University, Giza, Egypt in 2010.
- Experience:
  - Worked as graduate teaching assistant at the Department of Statistics, School of Economics and Political Science, Cairo University between September 2010 and August 2012.
  - Employed as graduate teaching assistant and instructor by the Department of Statistics, Oklahoma State University between August 2012 to July 2016.
- Professional Memberships:
  - Member of the American Statistical Association since January 2014.